THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVIII September, 1937 Number 6

THE DETERMINATION OF RESPONSE PATTERNS IN PERSONALITY ADJUSTMENT INVENTORIES*

JOSEPH ZUBIN

College of the City of New York

The widespread use of the personality adjustment inventories has recently evoked considerable criticism. Most of this criticism, however, has been directed not against the method itself but against its uncritical application. The method of self-description and self-appraisal that the inventory utilizes yields fundamental knowledge about an individual which, if interpreted properly, could be of service to the research worker or clinician. But in using this method one must guard against two major sources of error. The first type of error may result from the acceptance of the responses to the inventory at their face value, interpreting them logically rather than psychologically. The interviewer does not experience this difficulty to the same degree because he judges the personality of the patient not only by the verbal responses given to the interview questions, but also by the general context of these responses and the accompanying indications of affect. Whenever some doubt arises as to the patient's interpretation of the question, or as to the exact meaning of his response, the interviewer can clear up the issue by asking further questions. The method of administering the inventory does not provide any helpful cues for determining the meaning and veracity of the individual's response. But some statistical techniques are available which circumvent some of these difficulties, and they have been discussed in recent publications.[1,2]

The second source of difficulty inheres in the method used to obtain and interpret the total score for all the items in the inventory. It is this second source of difficulty that will be dealt with in this paper.

* This paper represents in part a summary of two papers read before the American Psychological Association, one in 1934 at the Columbia University meeting[10] and the second in 1936 at the Dartmouth College meeting.[9]

In most personality inventories every response to an item is weighted on some logical or empirical basis. The numerical values of these weights lie on a continuum ranging from the weight attached to the most adjusted response to the weight attached to the most neurotic response. The sum of the weights earned on the responses is regarded as the total score. That this total score is not very meaningful has not always been recognized. Two individuals who obtain identical scores on a personality inventory of the self-descriptive type may have earned this score by dint of ascribing to themselves entirely different sets of personality traits. There are, for instance, more than seventeen trillion different possible combinations, each yielding a score of ten in a test consisting of one hundred items. Yet all individuals who make a score of ten on a personality inventory are generally regarded as possessing equivalent degrees of personality adjustment. This practice is probably an uncritical application of the method used in obtaining total scores in intelligence tests. There, the assumption that a person who responds correctly to one set of items is equal in ability to a person who responds correctly to a set of ten other items is perhaps justified. Even in the case of intelligence tests, psychologists in general, and psychologists of the Gestalt school in particular, might contend that a knowledge of the pattern of test items passed or failed by a given individual may be more important than the knowledge of the total number correctly answered. This contention is even more pertinent in the interpretation of scores on personality inventories. Here the equivalence of the items has never been established and their additive character never proved. A total score that represents the summation of discrete responses to qualitatively different personality inventory items is, therefore, of little diagnostic significance. It is not surprising then to find that the total scores on some personality inventories fail to differentiate between groups that are so widely different in degree of adjustment as patients in a mental hospital and normal individuals. What can be done to obtain an index that will represent the cumulative effect of the responses of the individual and be more adequate than the usual total score? One solution is to find some method of representing the *total response pattern* of the individual. For a three-choice test we can let *Yes* be represented by 1, *No* by 2, and *?* by 3, and write down the pattern for the responses to, say, a five-item test as follows: If the responses are *Yes* to item 1, *No* to item 2, *?* to item 3, *Yes* to item 4 and *?* to item 5, the code number representing these responses is 12313. If the number of items in the test is increased to, let us say

seventy, a seventy-digit number would be required to represent this total response pattern, in which the digits could vary only from one to three.[6]

But this total pattern is at best a qualitative index and can not yield norms for the comparison of different individuals. The total response pattern is so individualized that very few individuals whose total response patterns are identical can be found. However, many individuals do give identical partial patterns, that is, their responses are identical for certain groups of items, though not for all.

The purpose of this study was to determine those items in which responses of individuals tend to group themselves into definite patterns. The first step was to administer nine different questionnaires containing a total of six hundred thirty-two items to a group of patients at the New York State Psychiatric Institute* and to an equal number of matched normal individuals.[1] The variables on which the individuals in the two groups were paired are age, intelligence, schooling, occupation, racial extraction, religion, and nativity. After analyzing each item to determine its power to differentiate between the clinical psychiatric categories and their normal controls, the items that proved to be diagnostic were arranged in a theoretical scale.

The next step in this study was to determine what patterns, if any, existed in these responses. The usual procedure for finding the items that belong together in one pattern is to determine whether any central or group factors exist in the items by applying a factor analysis utilizing either the Spearman,[5] Hotelling[3] or Thurstone[4] methods. First the tetrachoric correlation coefficients between the responses to the individual items are computed and then the factor analysis procedure is applied. Those items that are found to have common factors may be regarded as belonging to a pattern or constellation. The use of this method, however, involves several assumptions that are not always justified. For the tetrachoric correlations it is necessary to assume that the two response choices to each dichotomous item ("Yes" or "No") form a continuum and that the distribution of individuals in that continuum is normal. If there are three choices to each item (Yes, No and ?) it would be necessary to invent a new method of obtaining the correlation coefficient from the six-fold table, instead of the usual

* The work reported in this paper was carried on in the Psychological Laboratory of the New York State Psychiatric Institute and Hospital. The author is indebted to Dr. Carney Landis *Director of the Laboratory* for making this study possible.

four-fold.* Furthermore, the factor analysis method leads only to a factorization of the responses into underlying independent components. The primary purpose of this research was to factorize not the questionnaire items, but to determine whether or not patterns or clusters of item-responses can be found which correspond in a fairly uniform fashion to the diagnostic categories to which those taking the questionnaires had been assigned.†

By patterns we mean a cluster or constellation of responses to a given group of items. If the frequency of individuals exhibiting the pattern is significantly greater than the frequency expected by chance, the pattern is established. If the frequency is not significantly greater than the chance expectancy, it is discarded.

The assumption underlying the present approach is probably very similar to that utilized by systematists like Kraepelin or Bleuler in their formulation of the present-day psychiatric categories, namely, observing what symptoms seem to occur together as a syndrome. The approach utilized in this study differs from the classical approach, first, in the fact that this study utilizes self-descriptive symptoms only, rather than the symptoms observed by the medical psychologist and, secondly, that it utilized statistical analysis instead of free observation.

There are a tremendous number of theoretically possible patterns of response which might be obtained from our data. For practical and exploratory reasons we have limited ourselves to combinations of three, four, five and six items. Although it would be interesting to attempt to find differential patterns of all possible combinations of this variety, it is still beyond the realm of practicality. Some logical and empirical selection had to be made. Only the items that proved to be diagnostic (differentiating between the normal and abnormal groups) were utilized. There were one hundred forty such items. If all possible combinations of three of these one hundred forty items were

---

* Perhaps each one of the three responses could be counterposed in turn against the combination of the two remaining responses.

† It should be noted that another possible approach to the problem is to utilize the factorial analysis as the first step and to determine the various "factor" scores for each individual and then subject these hypothetical "factor" scores to the pattern analysis that is developed in this paper. But factor analysis presupposes a considerable number of assumptions, as indicated above, while the pattern analysis method requires only the assumptions that underlie the derivation of χ^2, which are minimal. Furthermore, since the factors found by the factorial analysis are by definition orthogonal, it is questionable whether any significant patterns could be discovered among these factor scores.

used, we should have had to make 12,084,660 possible combinations. This, being true, we were driven to the selection of item combinations on a sampling basis. Our method of sampling involved three procedures: First, the selection of items which might logically seem to be associated; for example,

1. Becomes upset easily.
2. Considered by friends as a person of strong nerves.
3. Falls asleep easily and is not troubled by thoughts about something that has happened.

We shall designate this triad of items, for convenience, as Triad I.

Our second procedure involved the choice of items which had no apparent connection; for example,

1. Likes to have people tell him frankly what they think about him even if it hurts, as long as what they tell him is true.
2. Dislikes people who speak very fast.
3. Sometimes has dizzy spells.

A third procedure for selection involved the putting together of certain symptoms which are said to be characteristic of one or another of the psychiatric diagnostic categories; for example,

TABLE I.—RESPONSES TO INDIVIDUAL ITEMS IN TRIAD I

	Yes		No		0*	
	N	Per cent	*N*	Per cent	*N*	Per cent
Item 1.						
Normal	27	26.0	76	73.1	1	0.9
Abnormal	56	50.9	53	48.2	1	0.9
	83	38.8	129	60.3	2	0.9
Item 2.						
Normal	69	66.4	33	31.7	2	1.9
Abnormal	47	42.7	63	57.3	0	0.0
	116	54.2	96	44.9	2	0.9
Item 3.						
Normal	55	52.9	48	46.2	1	0.9
Abnormal	40	36.4	67	60.9	3	2.7
	95	44.4	115	53.7	4	1.9

* 0 = omitted.

1. Thinks it possible for other people to influence his actions.
2. Prefers an adventurous life to a quiet home life.
3. Daydreams a lot.

The $Y_1N_2Y_3$ response to this triad would be accepted by many psychiatrists as characteristic of the Schizophrenic group.

Table I shows the distribution of responses to each of the items in Triad I.

In item 1, *No* is the normal response, being given by 73.1 per cent of the normals and only by 48.2 per cent of the abnormals. Similarly *Yes* is the abnormal response being given by 50.9 per cent of the abnormals and only by 26.0 per cent of the normals. The discrepancy between the proportion of the *Yes* and the *No* responses in the normal and abnormal group is too large to be explained as a chance discrepancy. The value of χ^2 for this comparison is 14.08 which for two degrees of freedom could arise by chance less than one time in one hundred.[7]

In item 2 *Yes* is the normal response with a frequency of 66.3 per cent in the normal group and of only 42.7 per cent in the abnormal group. *No* is the abnormal response with a frequency of 57.3 per cent in the abnormal group and of only 31.7 per cent in the normal group. The value of χ^2 for this comparison is 15.41 which could arise by chance less than one time in one hundred.

In item 3 the normal response is *Yes* and the abnormal *No*. The value of χ^2 for item 3 is 6.34 which could arise by chance less than one time in one hundred.

From a survey of the responses in Table I it might be concluded that the prevalent abnormal pattern of response should be *Yes* to item 1, *Yes* to item 2, and *No* to item 3, symbolically represented as $Y_1Y_2N_3$. Similarly the most prevalent normal response should be $N_1Y_2Y_3$. Table II shows the actual distribution of response patterns in the normal and abnormal group. The omissions are indicated by *Q*.

TABLE II.—RESPONSE PATTERNS TO TRIAD I

	NNN	*NNY*	*NYN*	*NYY*	*YNN*	*YNY*	*YYN*	*YYY*	*OYO*	*YNO*	*ONN*	*NNO*	*NOY*	Total
Normal....	10	8	23	32	8	5	6	8	0	0	1	1	2	104
Abnormal..	11	8	17	20	30	8	10	3	1	2	0	0	0	110

The expected normal pattern, $N_1Y_2N_3$, is given by thirty-two normals and twenty abnormals. The expected abnormal response *YNN* is given by eight normals and thirty abnormals.

We must now determine whether or not the discrepancy between the incidence of these patterns in the normal and abnormal groups is significant.

Reconstructing the twenty-six-fold table (Table II) into a six-fold table, we obtain Table III which gives the incidence of the *YNN* (expected abnormal pattern), *NYY* (expected normal pattern) and the remaining patterns grouped into one category.*

TABLE III.—INCIDENCE OF RESPONSE PATTERNS TO TRIAD I COMPARED WITH CHANCE EXPECTANCY COMPUTED FROM MARGINAL FREQUENCIES

	YNN			*NYY*			Other patterns			Total	
	*O**	*E*†	χ^2	*O*	*E*	χ^2	*O*	*E*	χ^2	*N*	χ^2
Normal.....	8	18.5	5.9351	32	25.3	1.7743	64	60.3	0.2270	104	7.9364
Abnormal.....	30	19.5	5.6129	20	26.7	1.6813	60	63.7	0.2149	110	7.5091
	38	38.0	11.5480	52	52.0	3.4556	124	124.0	0.4419	214	15.4455

* *O*—obtained.

† *E*—expected by chance.

From a consideration of the marginal frequencies in Table III it can be seen that by chance we should expect 18.5 individuals to give the *YNN* response in the Normal Group. The obtained value in the Normal Group was only eight. Similarly the obtained value in the abnormal group was thirty compared to a chance expectancy of only 19.5. The χ^2 value for this discrepancy between the number of normal and abnormal individuals giving the *YNN* response is 11.5480 which for one degree of freedom could arise by chance less than one time in one hundred. Apparently the *YNN* response is definitely associated with abnormality and any one who gives such a response is, other things being equal, more likely to belong to the abnormal than to the normal group. A similar analysis for the *NYY* pattern would indicate that it characterizes the normal group, though the probability of this characterization is equivalent to a χ^2 of 3.4556 which would arise by chance about eight times in one hundred.

But the above comparison considers the frequency of pattern responses such as *YNN* or *NYY* as the original datum. Actually, the individuals taking these tests did not respond to patterns of items,

* Table II could have been analyzed as it stands according to all the twenty-six-fold classifications, but the incidence of some of the patterns is so small that they were combined under the term, other patterns.

but to single individual items. Accordingly, instead of calculating the frequencies expected by chance from the marginal frequencies in Table III, we should calculate them from the frequency of responses to the individual items as given in Table I.

If there were no association between the individual responses, the incidence of the pattern $Y_1N_2N_3$ should be equal to the product of the incidences of the various responses included in this pattern, that is, the incidence of *YNN* responses in the normal group, should be the product (.388)(.449)(.537)(104) = 9.7.

We can now construct a table of values expected by chance for the response patterns. This is given in Table IV.

TABLE IV.—INCIDENCE OF THE RESPONSE PATTERNS AND THEIR EXPECTED CHANCE VALUES COMPUTED FROM INCIDENCE OF INDIVIDUAL RESPONSES

	YNN			*NYY*			Other patterns			Total	
	*O**	*E*†	χ^2	*O*	*E*	χ^2	*O*	*E*	χ^2	*N*	χ^2
Normal.....	8	9.7	0.2979	32	15.1	18.9145	64	79.2	2.9172	104	23.1296
Abnormal.....	30	10.3	37.6786	20	15.8	1.1165	60	83.9	6.8082	110	45.6033
	38	19.0	37.9765	52	30.9	20.0310	124	164.1	9.7254	214	68.7329

* *O*—obtained.

† *E*—expected by chance.

The χ^2 value for this comparison is 68.7329. Since the expected values were not obtained from the marginal frequencies there are four degrees of freedom in this table, for any four cells could be filled at random without necessitating any change in the two fixed marginal frequencies—the frequencies in the normal and abnormal groups.[7] A value of χ^2 as large as the one obtained could arise by chance less than one time in one hundred. This table could not have arisen by chance. What patterns have yielded frequencies significantly in excess of chance? In the *YNN* pattern, the normal group yielded a frequency of eight and a chance expectancy of 9.7. This discrepancy is so small that all the normals who have given this response pattern can be regarded as having given it by chance rather than by choice. In the abnormal group, however, we expected only 10.3 individuals and obtained thirty. This discrepancy can not be explained by chance. Reconstructing this table so as to compare the frequency of the *YNN* pattern of response with the frequency of all other response patterns in the normal

and abnormal group, we obtain Table V which yields a χ^2 of 41.8997 that can occur by chance less than one time in one hundred.

TABLE V.—OBTAINED AND EXPECTED VALUES FOR ABNORMAL RESPONSE PATTERN TO TRIAD I

N	*YNN*			*NON – YNN*			Total	
	O	*E*	χ^2	*O*	*E*	χ^2	*N*	χ^2
Normal.....	8	9.7	0.2979	96	94.3	0.0306	104	0.3285
Abnormal.....	30	10.3	37.6786	80	99.7	3.8926	110	41.5712
	38	20.0	37.9765	176	194.0	3.9232	214	41.8997

There seems to be little doubt that the *YNN* response is characteristic of the abnormal group.

It should be noted that the method of determining the value expected by chance in Table V seems to harmonize more with the hypothesis that *YNN* is an abnormal response than the method utilized in Table III. In Table V it is quite apparent that the eight individuals of the normal group who gave the abnormal response may quite properly be ascribed to chance since the number expected by chance is 9.7 and the obtained value is eight. The discrepancy is truly a random one. In Table III the number expected by chance is eighteen which is ten points larger than the obtained value. This discrepancy gives rise to a χ^2 of 5.9, a value which approaches significance.

Turning now to the *NYY* pattern we note that in the abnormal group the frequency expected by chance was 15.8 and the frequency obtained was twenty, an excess of 4.2 cases which yields a χ^2 of 1.1165. In the normal group 15.1 cases were expected by chance and the number obtained was thirty-two, an excess of 16.1 cases yielding a χ^2 value

TABLE VI.—OBTAINED AND EXPECTED VALUES FOR NORMAL RESPONSE PATTERN TO TRIAD I

	NYY			*NON – NYY*			Total	
	O	*E*	χ^2	*O*	*E*	χ^2	*N*	χ^2
Normal.....	32	15.1	18.9145	72	88.9	3.2127	104	22.1272
Abnormal.....	20	15.8	1.1166	90	93.2	0.1893	110	1.3058
	52	30.9	20.0310	162	182.1	3.4020	214	23.4330

of 18.9145. Reconstructing Table IV into a four-fold table with the *NYY* response as the pivotal column we obtain Table VI:

The χ^2 value for this table is 23.4330 which could arise by chance less than one time in one hundred. The *NYY* response is definitely associated with normality although the probability of this association is less than that of the *YNN* response with abnormality. The remaining patterns do not seem to bear any association to the normal group, but some of them are apparently associated with the abnormal group as indicated by the marginal χ^2 of 9.7254 (Table IV).

This analysis was applied to all the patterns that had been selected for investigation. Although each item that entered any of the triads was itself a diagnostic item, differentiating individually between normal and abnormal groups, not all the triads of these individually diagnostic responses proved to be diagnostic as a pattern. The number of normal and abnormal patterns in the responses of each individual were computed and the averages by clinical categories are given in Table VI.

TABLE VI.—AVERAGE NUMBER OF NORMAL AND ABNORMAL PATTERNS BY CATEGORIES

Category	*N*	Normal patterns	Abnormal patterns	Difference
Schizophrenic	37	9.24	3.27	5.97
Manic depressive	32	10.81	2.56	8.25
Psychoneurotic	25	6.80	4.12	2.68
Organic	15	13.20	1.73	11.47
Total abnormal	109	9.69	3.05	6.64
Total normal	104	14.35	1.08	13.27

The normal group averaged 14.35 normal triads and only 1.08 abnormal triads. The abnormal group averaged 9.69 normal triads and 3.05 abnormal triads. Apparently, both the normal and the abnormal groups show considerable pattern integration. The normals, however, yield more integrations than the abnormals.

Taking the difference between the number of normal and the number of abnormal triads as a tentative measure of normality, we note that the normal group has a score of 13.27, compared to a score of 6.64 for the abnormal group. A comparison of the scores by categories shows that the psychoneurotic group made the lowest score and the

manic-depressive the highest score. The paretics scored almost as high as the normals.

It might be objected that this process of adding together patterns to obtain a total pattern score is erring in the very direction that this study attempts to criticize. However, this summing up is used only for group comparisons. In the case of a single individual it would not be sufficient to add up the number of abnormal patterns. For individual diagnosis a syndrome of patterns is to be utilized, if such syndromes are discovered.

The pattern scoring method eliminates another fault of the ordinary scoring methods. Most studies comparing normal and abnormal individuals have found that there is considerable overlap in total scores between the normal and abnormal individuals. An analysis of Table I indicates that this overlap is a necessary result of using the total score as an index of adjustment. In Table I we note that although the *No* response is the *normal* response to item 1, fully 48.2 per cent of the *abnormal* group give that response. Similarly 42.7 per cent of the *abnormal* group gave the *normal* response to item 2 and 36.4 per cent of the *abnormal* group gave the *normal* response to item 3. Small wonder then that there is much overlap between the two groups. By utilizing the pattern score the overlap is reduced considerably. Thus, as shown in Tables III and IV, only eight individuals or 7.2 per cent of the *normal* group are found to give the *abnormal* response pattern and twenty individuals or eighteen per cent of the *abnormal* group gave the *normal* response.

The major difficulty to be overcome before pattern scoring can replace item scoring is the development of a simple scoring device to enable the determination of the patterns that occur in a given set of responses. Such a scheme has already been devised and consists in the main of a technique for coding the item responses in such a manner that the code numbers yield a unique index of the pattern.* Another technique utilizing the punched card method is now being developed by the writer in collaboration with T. W. Forbes.[9] The work of finding the index will eventually involve no more labor than the ordinary work involved in scoring the usual test. After the pattern

* Toops[8] has suggested a scheme which serves our purpose when the items are treated in small groups. The present writer has evolved a scheme involving the use of the products of prime numbers which is for our purposes more serviceable. This new scheme together with the logarithmic codes of the prime numbers will be published shortly.

index has been obtained, the value of the index can be determined from tables that have been prepared for this purpose.

The findings of this study lead to interesting speculations concerning the problem of the organization of the mental life of the psychopathic patient. It is a moot question whether insanity is characterized by the total breakdown of the normal patterns of response involved in thought or whether it is characterized by the establishment of new, queer, and abnormal patterns of response. The results of this study indicate the possibility that both processes may take place. The abnormals show fewer of the normal patterns and also exhibit unique patterns that are characteristic of their own group. But it is possible that the results obtained in our study are partly obscured by the admixture of patients of different degree of severity of illness.

In addition to the triad analysis reported above, similar analyses were made of patterns involving four, five, six, ten and twelve items. The results of one such analysis are shown in Table VII.

TABLE VII.—DISTRIBUTION OF RESPONSE PATTERNS TO A HEXAD

	1 *NNYNYY*			2 *YYNYNN*		
	Obtained	Expected	χ^2	Obtained	Expected	χ^2
Normal.....	21	6.77	29.91	0	0.12	0.12
Abnormal.....	12	7.10	3.38	5	0.13	16.*

* Counting the expected value as one.

The first pattern is characteristic of normal individuals and the second of abnormal individuals. A general survey of these higher pattern analyses reveals the fact that there is a diminishing return in number of significant patterns as the number of items in the pattern increases. The optimum number of items to a pattern seems to be five. This result must be presented, however, as tentative, since it depends most probably upon the size of the total group as well as on the method of selecting the pattern.

SUMMARY

The purpose of this study was to determine whether or not patterns exist in the responses of normal and abnormal subjects to personality adjustment inventories. It was found that such patterns do exist and that the scoring of such tests by patterns rather than by individual

responses is possible. Normal individuals possess more integration as revealed by the greater number of patterns that they exhibit. Abnormal individuals exhibit some patterns peculiar to themselves, but also exhibit some of the normal patterns. There appear to be differences between the psychiatric categories in this respect. The psychoneurotics exhibit more abnormal patterns and fewer normal patterns than do any of the other categories. On groups of the size utilized in this study, two hundred, the optimum size of pattern for analysis is the triad. This method still requires certain mechanical improvements to facilitate the scoring. Until these methods are discovered, the pattern analysis described above will be of use only as a research tool. But it promises to be of use in determining personality patterns of both normal and abnormal individuals.

REFERENCES

1. Landis, C., Zubin, J., and Katz, S. E.: "Empirical Evaluation of Three Personality Adjustment Inventories." *Journal of Educational Psychology*, 1935, pp. 321–330.
2. Benton, Arthur L.: "The Interpretation of Questionnaire Items in a Personality Schedule." *Archives of Psychology*, 1935.
3. Hotelling, H.: "Analysis of a Complex of Statistical Variables into Principal Components." *Jour. Educ. Psych.*, Vol. XXIV, 1933, pp. 417–441.
4. Thurstone, L. L.: "The Vectors of Mind." *Psych. Rev.*, Vol. XLI, 1934, pp. 1–32.
5. Spearman, C.: *Abilities of Man.* Macmillan, 1927.
6. Zubin, J.: "The Coding of Item Responses for Pattern Analysis" (to be published).
7. Fisher, R. A.: *Statistical Methods for Research Workers.* Edinburgh: Oliver Boyd, 1934.
8. Toops, H. A.: "Questionnaire Construction and Analysis" in Baehne, G. W., *Editor, Practical Applications of the Punched Card Method in Colleges and Universities.* Columbia University Press, 1935.
9. Zubin, J.: "A Technique for Pattern Analysis." *Psychological Bulletin*, Vol. XXXIII, 1936, p. 733, abstract. Read before the meeting of the American Psychological Association at Dartmouth College, 1936.
10. Zubin, J.: "The Determination of Respose Patterns in Personality Inventories." *Psychological Bulletin*, Vol. XXXI, 1934, p. 712, abstract.

THE OPTIMAL TEACHING SEQUENCE FOR ELEMENTARY PHYSICAL PRINCIPLES BASED ON A COMPOSITE SCALE OF PLEASURE-VALUE AND DIFFICULTY OF INSIGHT

GEORGE W. HARTMANN AND DEAN T. STEPHENS*

Teachers College, Columbia University

IS THERE ONE BEST ORDER OF PRESENTATION IN DEVELOPING MASTERY OF SCIENTIFIC CONCEPTS?

Logicians, psychologists, and teachers of special technical subjects have long known that the skills and comprehensions required for the effective learning of certain fields of knowledge can be ordered in some sort of hierarchical pattern. If we may judge by the errors pupils make with different number-combinations, it is much harder to learn that $9 + 8 = 17$ than that $2 + 2 = 4$. Some basic logical relations of broad range are apparently easier to grasp than others, for children give evidence of sensing the "form" of the *action-agent* situation before they are able to handle the *species-genus* grouping. Similarly, there is good empirical reason for believing that the minds of secondary-school students are so organized that they easily see how "Gases tend to expand and completely fill their containers" but fail relatively to assimilate the fact that "When an e.m.f. with the same frequency as the natural frequency of a circuit is applied to that circuit, electrical resonance is produced."

Variations in the "altitude" of intelligence demanded for the comprehension of these principles is probably the major psychological limitation involved in the examples just cited, but familiarity with the concrete phenomena underlying the generalizations and the accompanying affective experience of the "value" of such knowledge also facilitate or inhibit the understanding, particularly at maturer stages of learning. In well-systematized areas of instruction, such as introductory physics, a pure or "intrinsic" difficulty scale of major concepts could undoubtedly be established. However, not all stimuli are perceived as of equal worth to the learner. The structure of physical thought is not homogeneous in interest-merit at all its levels and the consequent fluctuations in enthusiasm influence the apparent ease of its different sections. This is recognized as affecting preference among

* This study was begun while the authors were associated with the School of Education of the Pennsylvania State College.

the major "subjects" of instruction, but is often neglected so far as the internal organization of the content of each discipline is concerned.

We have reason to believe that the teaching of physics can be made more successful as a scholarly acquisition as well as a vivid communication of the "excitement" of agreeable progress if its key principles fit the specific growth stages of the learner. Blake's *Natural Philosophy*, published in 1835, shows approximately the same organization as the physics textbooks of today, with the exception that astronomy, then included in the course, has become a separate field of instruction. It is undoubtedly true that the principles of the physics of sound, light, and heat are fairly "natural units" of instruction in the general area of physical science as such, but it would be most surprising if the order in which they appeared were a matter of indifference. It is a grotesque inversion of the logic of temporal succession to grapple with quadratics before facility in the operations of factoring and transposition have been acquired. In less well-integrated subject-matter fields, such a displacement may be tolerated, but in areas that have achieved a certain formal coherence it necessarily results in lowered instructional efficiency. The present paper is an effort to supplant the judgments of experts and the unformulated hunches of experienced teachers with detailed evidence for an order of exposition adapted to the developmental peculiarities of the average student of high-school physics.[1]

SELECTING THE PRINCIPLES FOR RE-ORGANIZATION

The gist of elementary physics is assumed to lie in the most widely-used textbooks of that subject. Publishers' sales reports were used as indices of popularity. The six commonest high-school texts at the time this study was begun are listed below.[2] Careful underlining of the

[1] The practical pedagogical value of such inquiries may be seen in a related study by A. B. Lewis. He prepared an algebra sequence according to the progression of sheer "logical inference" and taught paired classes with this "curriculum" and the sequence of the Philadelphia course of study. Common tests given at the end of each term favored the new organization. See *A scientific topic sequence of the subject-matter in the elements of algebra*, unpublished master's thesis, Temple University, 1928.

[2] Black, Newton H. and Davis, Harvey N.: *New Practical Physics*. New York: Macmillan, 1930.

Brownlee, Raymond B. and Others: *First Course in Physics*. New York: Allyn & Bacon, 1930.

Dull, Charles E.: *Modern Physics*. New York: Henry Holt, 1929.

Henderson, William D.: *New Physics for Everyday Life*. New York: Lyons & Carnahan, 1930.

content of each book led to the segregation of all discriminable principles, laws, corollaries, and "key" phenomena. Definitions of terms, illustrations, and detailed facts were usually amplifications of some readily-identifiable principle. Somewhat to our surprise, we found that the six volumes together employed no more than two hundred sixty-three separate principles; the number of "concepts" was considerably greater, of course, but a principle was conceived as a functional relation or "whole" involving the interconnection of two or more concepts. The psychological and semantic division between *concept* and *principle* is not a sharp one save that (in physics at least) the relational and implicitly mathematical features are more prominent in the latter.

The two hundred sixty-three propositions originally identified were eventually reduced to one hundred forty-four by eliminating every item that did not appear in *all* the text sources (a severe reduction to minimum essentials) and all those that seemed like duplications or closely related statements in the judgment of four physicists. It is interesting that the representation given to the conventional subdivi-

TABLE I.—DISTRIBUTION OF PHYSICAL PRINCIPLES AMONG THE DIFFERENT BRANCHES OF THE SCIENCE ACCORDING TO TWO INDEPENDENT TABULATIONS

(1) Conventional grouping	(2) Number of items in present check list	(3) Column (2) in percentage terms	(4) Proportion used by Columbia Research Bureau Physics Test[1]
Mechanics	38	26.38	16
Heat	19	13.20	16
Electricity	50	34.72	32
Sound	8	5.56	8
Light	29	20.14	16
Miscellaneous	0	0.00	12
Total	144	100.00	100

[1] From Hermon Farwell and Ben D. Wood, Yonkers, World Book Co., 1926. Most of the items classified as "miscellaneous" here fall under "mechanics" in our list. This makes the two distributions coincide even more closely.

Holley, Clifford and Lohr, Vergil C.: *Mastery Units in Physics.* Philadelphia: Lippincott, 1932.

Stewart, Oscar M. and Others: *Physics for Secondary Schools.* New York: Ginn, 1932.

sions of physics follows closely the percentages used in the final edition of the Columbia Research Bureau Physics Test where the content occurs in problem form. Presumably independent "samplings" from the same frame of reference, even when made a decade apart with different sources, must yield approximately the same results.

MEASURING THE RELATIVE DIFFICULTY OF EACH PRINCIPLE

The degree of resistance to the emergence of an insight is here used as the most meaningful measure of difficulty in learning. Insight into a principle is considered lacking if a learner makes one hundred per cent errors in his attempted solutions; it is considered present if one hundred per cent successful solutions follow its advent. This rigorous criterion is hard to apply because of the existence of partial, gradual, and imperfectly developed or recalled insights,[1] but it seemed the best one to use in following an expanding learning situation. Twelve students with IQ's ranging from one hundred twenty to one hundred forty-two were examined regularly, individually and intensively while they were members of the junior author's class in high-school physics during the year 1935–1936. The order of difficulty established for superior students is probably the same for average and inferior students, even though the general level of the series is displaced upward or downward.

Two days after a principle had been studied or expounded, five "application" questions concerning it were administered. Thus, the generalization that "When a body is heated the average speed of molecular agitation increases" (Number 10 in the check-list) was followed by these problems:

(*a*) Why does a body expand when heated?
(*b*) Why does hot water evaporate more rapidly than cold?
(*c*) Why do metals become hot when pounded?
(*d*) What is the relation between temperature and the molecular theory?
(*e*) Why are gases, having fast moving molecules, poorer conductors of heat than metals?

The seven hundred twenty questions employed (144×5) were selected with as much care as the basic principles, after a preliminary trial during the preceding school year with a different but comparable group. Questions that none answered correctly were excluded from the main inquiry, although for motivation purposes as well as because

[1] Cf. *Gestalt Psychology* (Ronald Press, 1936), Chapter 12, by George W. Hartmann for a theoretical and experimental analysis of these points.

of their undeniable basic importance a fair number of those that everyone got right were retained. Each principle received a "difficulty score" computed as follows: Suppose that out of twelve students involved, four gave correct responses to the first question, eleven to the second, ten to the third, nine to the fourth, and eleven to the fifth. This makes a total of forty-five correct answers out of a possible sixty, or a "difficulty index" of seventy-five per cent. The mean "group insight"—if one may quantify such all-or-none phenomena—was 88.72 per cent for the heat principles, 86.18 for mechanics, 83.21 for light, 80.10 for sound, and 78.56 for electricity.

It may be argued that our insight into principles can be made to seem good or poor by the nature of the selection of certain types of problem-situations that illustrate the operation of the principle with varying degrees of clarity. Thus, the group's insight into one principle may be made spuriously high by choosing five easy applications of it, while its performance with respect to another principle can be lowered merely by demanding the solution of troublesome "teasers." Adequate sampling might show the mean difficulty of the first principle to be objectively greater than that of the second, but the special "packing" of the questions could invert the order. This difficulty with respect to the questions was recognized and combatted by "graduating" them so that some easy, some moderate, and some hard solutions were demanded in connection with each principle. Practically all the problems were derived from recent text sources and represent carefully considered and formulated questions. The arithmetical ability involved in their solution was kept at a minimum so that different degrees of mathematical competence would not affect the apparent understanding of a basic physical principle. Instructional emphasis was kept as nearly constant and uniform for all parts of the course as the mild oscillations in the classroom teacher's psychophysical tone allowed.

MEASURING THE RELATIVE PLEASURE VALUE OF EACH PRINCIPLE

The second investigational task was to measure the degree of interest associated with each principle. The procedure followed the pattern described in an earlier study by the senior author with improvements derived from experience with related inquiries.[1] The one hundred forty-four principles were staggered in a printed booklet, care being

[1] See "Variations in affective tone of different areas of educational psychology." *J. Ed. Psychol.*, Vol. XXV, 1934, pp. 115–135.

taken to see that a large number of items belonging to a single division of physics did not appear together. This was designed to eliminate extreme responses caused by the "total affect" of a larger field; thus, it is known that some pupils dislike *mechanics* as a whole but react favorably to the section on *optics*. Each principle was to be judged by itself, insofar as this isolation could be achieved by the respondent. The "interest" document when used by the subjects looked like this:

Below are a number of items which are found in a high-school physics course. You will readily recognize that these items do not have the same value or meaning to you. It is to ascertain these differences that this study is being made.

Kindly check the items according to the *degree of pleasure; i.e.*, the degree of satisfaction, agreeable emotion, or feeling, that it gives you to understand the item when you meet it in your experience.

If the knowledge of the item gives you a *feeling of aversion; i.e.*, a sense of annoyance, displeasure, or dissatisfaction, indicate the extent in the appropriate column.

If, however, you are *indifferent* or neutral in your feelings toward the subject-matter of the item, in that it causes you neither pleasure nor aversion, place a check in the column marked *indifferent*.

AVERSION				PLEASURE		
Extreme	Moderate	Slight	*Neutral or Indifferent*	Slight	Moderate	Extreme

If there are some items which you have never learned, do NOT attempt to check them.

Check Each Item According to Pleasure, Aversion, or Indifference

1. The law of the lever states that the effort times its distance from the fulcrum is equal to the resistance times its distance from the fulcrum.
2. &c.

This test was taken by two hundred fifty-two students who had *just finished* a year of elementary physics. Complete records were obtained from thirty-nine pupils in Brookville, Pa., one hundred fifty-one in East Orange, N. J., twenty-nine in Lock Haven, Pa., eleven in Mill Hall, Pa., nine in Scotch Plains, N. J., and thirteen in Springfield, Pa. One hundred seven of this group were junior boys, seventy-one senior boys, fifty-eight junior girls, and sixteen senior girls. Since

more boys than girls commonly elect physics, the sex ratio does not seem unbalanced.

Of the two hundred fifty-two students who answered the "check-list," twenty-one began their judgments at item Number 1, twenty-one at item Number 13, etc.; the place of beginning was indicated by a red mark. This scheme was adopted in order to prevent a possible reduction in reliability caused by a concentration toward the end of a number of unfinished items. Any stereotypy of response is also minimized by this method of distributing the fatigue factor.

The interest blanks were scored by assigning a numerical equivalent to each of the seven columns, thus:

Pleasure.	
Extreme	3
Moderate	2
Slight	1
Neutral or indifferent	0
Aversion.	
Slight	1
Moderate	2
Extreme	3

The mean pleasure value of each principle was calculated readily by summing algebraically its rating in all papers and dividing that total by the number of persons answering the item. The resulting quotient presumably measures the amount of "kick" or zest felt by the group in reacting to the emotional overtones of the statement.

The mean pleasure value for all papers was .94 $\pm$ .035, with an SD of .42. Happily, none of the items yielded a negative pleasure index. The highest average appeal is found in the principle, "Unlike poles of magnets attract each other; like poles repel each other." This has a pleasure value of 1.931, equivalent to "moderate." Perhaps its "human" and literary uses as well as its aesthetic simplicity account for this preëminence. The lowest value attaches to the statement, "Electrostatic induction is the separation of charges on a conductor through the influence of a neighboring charge," which approached indifference with a value of 0.064. Evidently the field theory of configurationism is not a source of excitement to high-school students.

The descending order of the mean pleasure indices for the different physical divisions ran as follows: Light 1.101, mechanics 1.054, heat 1.007, sound 1.004, and electricity .718. Save for the low status of electrical principles, the hedonic tone of the other divisions is surprisingly alike. These group values tend to erase the serial position of the separate principles, but this artifact should not blind one to the

substantial correlation between the mean pleasure value of each item and its difficulty (or ease) of insight as measured by the per cent correct in the five-problem situation described above. The Pearson product-moment r's between these two quantities are .66 for light, .56 for mechanics, .75 for heat, .74 for sound, and .68 for electricity. This is probably the familiar correlation between interest and ability appearing again in a subtle disguise.

The consistency of the "interest" test was determined by pairing scores from two sets of seventy-five papers each, selected by taking a boy's and then a girl's paper from each school in rotation. From these two groups, two independent sets of "mean" scores were obtained for each item. The resulting correlation was .93 $\pm$ .01 (where $N = 144$). The *grand* mean pleasure value of the seventy-two odd items for each of the two hundred fifty-two subjects was also correlated with the corresponding quantity for the seventy-two even items, yielding an r of .94 $\pm$ .01 (where $N = 252$). These two coefficients suggest satisfactory reliability for the Pleasure-Aversion test.

There were noticeable differences in the mean pleasure indices in the population of the various schools. These values were: Lock Haven, 1.603; Brookville, 1.161; Springfield, 1.137; East Orange, .785; Mill Hall, .680; and Scotch Plains, .508. These differences reach or exceed the conventional thresholds of significance only for the more extreme values. It is not clear why the attractiveness of physics should fluctuate thus widely from one community to another, but this alteration in general level does not necessarily detract from the ability of the interest check-list to reveal the location of the specific principles in the pleasure hierarchy. Even with the small samples from Lock Haven and Brookville the correlation between these school systems' pleasure values for each item was .54.

Analyzing a sample of fifty boys' and fifty girls' papers, we found that the mean pleasure index for the former was 1.15, SD .65; for the latter, .82, SD .58. The chances turn out to be 99.63 in one hundred that the male value is higher than the female, and support the common claim that in our culture physics appeals more to one sex than to the other.

It is also of interest that statements of more than average word number (in this case, the mean was 21.2) are lower in pleasure value than the short-sentence principles. The chances run 98.52 in one hundred. Inspection indicates that the long statements tend to be more "technical" than the short, although it should not be forgotten that they are "long" largely because of their technicality.

CONSTRUCTION OF A COMBINED LEARNING SCALE FOR THE PRINCIPLES USING BOTH PLEASURE AND DIFFICULTY VALUES

Can we fruitfully unite the paired notions of an ascending order in difficulty of insight and a descending scale of hedonic tone into a single but complex measure of "physics readiness"? A pupil's readiness to learn is clearly a function of at least these two forms of intellectual and emotional maturity. Strictly speaking, one does not have physics readiness as such, or even mechanics readiness, but such a circumscribed "set" as readiness for principle Number 119 to be learned before principle Number 120.

There seem to be no serious theoretical objections to giving equal weight to the interest and difficulty factor. Perhaps difficulty should be given somewhat greater weight, since "altitude of intellect" is largely a fixed and irreplaceable attribute, whereas the "pleasure values" fluctuate with the differential application of extrinsic stimuli and motives. Interest seems to be more often a function of ability than the converse, although below the plane of absolute incapacity, the reverse situation is a common occurrence. The practical ease of equivalent weighting, however, plus its simple and defensible operational meaning made it the preferred procedure. Each scale can be adjusted to have an upper limit of one hundred (the difficulty measure, beginning as a percentage of correct solutions had this peak at the outset) by considering the maximum pleasure value of 1.931 as one hundred and multiplying each pleasure index by 51.78 (*i.e.*, 100 ÷ 1.931). Consequently, by adding the *product* of the pleasure index and 51.78 to the degree of insight attained, we derive for each principle a third value which we shall call the "combined index."

In the Appendix (which is really the chief table of this study) will be found all the important psychological "constants" for the one hundred forty-four principles of introductory physics. The items themselves are now finally arranged in accordance with their composite pleasure-insight values. The Appendix to some degree brings order out of a chaos deliberately produced at the start, for it reports the significant data not in their original "scrambled" but in their finished and organized form. In order to avoid unwieldy and high numbers designating serial position, the rankings for each of the conventional subdivisions of physics are presented separately.

The mean combined index for the principles of heat is 140.91, for mechanics 140.78, for light 139.93, for sound 132.10, and for electricity 115.74. The only completely reliable difference is between the electric-

ity section and the others, although the "chances" of a significant difference for the sound category (in all pairs other than electricity) run about ninety-seven in one hundred. Perhaps the relative "intangibility" of sound, and particularly electrical, phenomena accounts for this arrangement. Perhaps, too, the notion of "field forces" which may well be basic to our understanding of nature, has not been made to play a sufficiently prominent rôle in conventional physics instruction.

It is significant that almost all physics texts place mechanics first and heat second. To that extent, practice may be said to have anticipated the results of research. However, it rarely happens that the other three divisions are presented in the sequence that we feel disposed to recommend, *viz.*, light third, sound fourth, and electricity fifth or last.

CONCLUDING REMARKS

Within these broad areas, certain kinds of principles obviously show a psychological as well as a logical or methodological coherence. A resourceful physics teacher who is not limited by a fixed course of study should find the Appendix rich in suggestions for grouping "seminal ideas." It provides him with the basis for introducing one set of concepts early in the sectional development, another toward the middle, and still another toward the end. It would probably be unwise literally to take the "combined index" for each principle at face value by teaching an easy and attractive principle of heat, then following that by its nearest equivalent from electricity, etc., before going on to the harder and less appealing ones. This possible procedure, however, should not be too readily dismissed as that is essentially what is done by the "spiral" practice of elementary, intermediate, and advanced courses in all sciences. Perhaps the teaching program should be so arranged that all the principles of more than median "pleasure-insight" indices come in the first half of the course and those with less than median rank come later. Interest certainly grows by what it feeds upon and the learner's level of pretense or aspiration is definitely conditioned by his success or failure with early problem-situations of a given class.

This may be the only sound biological way of achieving that architectonic integration so admired by the orderly mind of the ripe scholar. Organisms actively *create* mental unification in themselves; they do not simply absorb such a condition as a completed product. This natural process is facilitated and not hampered whenever teaching occurs so that concepts and principles reinforce each other. Proximity of place-

ment is one of the best ways of ensuring this result. Insofar as the "unit plan" makes sense, it requires the *organization* of material around a *series* of important understandings. This organization may be a fortuitous one or a systematic one, depending on the degree to which it harmonizes with the inner development of a growing mind. It is this type of *bio-logic* that the findings of this inquiry seek to express.

APPENDIX

This tabular material contains the heart of the present investigation. The "item number" of each principle indicates the order in which it was printed on the blank for determining the pleasure-value of the one hundred forty-four statements. The original measuring units and the conversion into rank-placement are given separately in appropriate columns for the composite "pleasure-insight" scale, the difficulty scale, and the "interest" scale.

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Mechanics							
Energy can never be created or destroyed; it can be changed from one form to another with exact equivalence	199.10	1	99.16	1	1.930	1	87
Gases tend to expand and completely fill their containers	178.85	2	98.33	2	1.555	4	39
Disregarding friction the work put into a machine is equal to the work gotten out of it	176.88	3	99.16	1	1.501	5	131
When a body is floating or is submerged in a liquid or gas the buoyant force on the body is equal to the weight of the liquid or gas displaced	172.37	4	87.50	14	1.639	3	134
When a strong wind blows across the top of a chimney or tube, some of the air is drawn out, creating a partial vacuum and draft in the tube	171.58	5	85.00	16	1.672	2	11
A body is stable if its center of gravity lies above its base	170.25	6	98.33	2	1.388	7	135
Neglecting air resistance a body will fall with a constant acceleration of thirty-two feet per second	167.89	7	92.50	8	1.456	6	58

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Work is measured by the product of the force exerted and the distance through which it moves	166.20	8	97.50	3	1.327	8	7
The law of the lever states that the effort times its distance from the fulcrum is equal to the resistance times its distance from the fulcrum	161.02	9	95.83	5	1.259	11	1
Pressure in a still liquid is the same at all points on the same level	156.06	10	95.00	6	1.179	14	103
A liquid tends to keep its volume constant; it may take any shape, but when at rest keeps its top surface level	155.85	11	89.16	12	1.288	9	2
When there is a gain in force by using a simple machine there is a loss in speed and *vice versa*	155.25	12	91.66	9	1.228	12	102
When a gas expands, the pressure it exerts decreases; when a gas is compressed, the pressure increases	153.85	13	90.83	10	1.217	13	83
The moment of a force is the product of the force by the length of the lever arm	149.73	14	95.83	5	1.041	19	51
The efficiency of a machine is equal to the ratio of the output to input	149.73	14	95.83	5	1.041	19	25
The resultant of any two forces acting at the same point may be represented by the diagonal of a parallelogram constructed with the forces as sides	148.89	15	92.50	8	1.089	16	55
A change in force by a simple machine is expressed as mechanical advantage and is equal to resistance divided by the effort	146.11	16	96.66	4	.955	21	23
When the resultant of all the forces acting on a body is zero the body will stay at rest if at rest, or it will keep in motion at a constant speed in a straight line if it is in motion	145.33	17	79.16	21	1.278	10	136
Pressure in liquids in the open is produced by the weight of the liquid and varies at different levels	145.13	18	89.16	12	1.081	17	52
Resistance times length of an inclined plane equals the force required to move the resistance times the height of the plane	142.26	19	94.16	7	.929	22	142

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
If the adhesion of a liquid to a solid is greater than the surface tension of the liquid, the liquid wets the solid and in tubes of the latter the liquid will rise	142.05	20	83.33	18	1.134	15	85
A single force may be resolved into two components by drawing a rectangle whose diagonal represents the original force	136.65	21	90.00	11	.901	25	84
Pressure applied to an inclosed liquid is transmitted equally in all directions and is exerted perpendicularly to the inclosing surface	135.20	22	81.66	20	1.034	20	120
The free surface of a liquid contracts to the smallest possible area producing surface tension	132.70	23	78.33	22	1.050	18	57
The distance a body travels, starting from rest with a constant acceleration is one-half the acceleration times the square of the time	130.34	24	82.50	19	.924	23	122
The speed of a body at constant acceleration is equal to the product of the acceleration by the time	128.39	25	86.66	15	.806	28	6
When one body exerts a force on a second body, the second body exerts an equal and opposite force on the first	121.91	26	75.00	25	.906	24	3
All the moments tending to rotate a bar in one direction are equal to the sum of all the moments tending to rotate it in the opposite direction when the bar is at rest	120.36	27	76.66	24	.844	26	82
When forces act in the same direction, the resultant is their algebraic sum	118.16	28	89.16	12	.560	35	26
Sliding friction is proportional to the mass and independent of size of surface	113.18	29	88.33	13	.480	37	28
The volume of a gas varies inversely as the pressure to which it is subjected, other things being constant	112.69	30	84.16	17	.551	36	104
Momentum is a measure of the quantity of motion that a body possesses; *i.e.*, mass times the velocity	112.15	31	70.00	28	.814	27	8
The amount of change in the shape of an elastic body is proportional to the force applied, provided that the elastic limit is not exceeded	110.69	32	75.50	23	.641	31	106

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
The energy which a body possesses on account of its motion is called kinetic energy and is proportional to its mass and the square of its speed	109.73	33	74.16	26	.687	30	32
The time of vibration of a pendulum does not depend upon the nature or weight of the bob, or on the length of the swing, but does vary directly as the square root of the length and inversely as the square root of the acceleration of gravity	106.94	34	65.83	31	.794	29	137
The acceleration of a body is proportional to the resultant force acting on that body and is in the direction of that force	103.92	35	71.66	27	.623	34	133
Centrifugal force is directly proportional to the speed squared, the mass, and inversely to the radius of rotation	101.57	36	69.16	29	.626	33	60
All bodies attract one another with a force which is directly proportional to the attracting masses and inversely proportional to the square of the distance between their centers of mass	100.59	37	67.50	30	.639	32	107
Heat							
While freezing the temperature of a liquid cannot become lower than its freezing point	177.49	1	95.83	3	1.577	2	90
Most bodies expand on heating and contract on cooling; the amount of change varies directly as the absolute temperature	174.35	2	90.00	10	1.629	1	61
Evaporation is a cooling process because heat is absorbed in changing a liquid into a vapor	167.22	3	94.16	4	1.411	3	144
Freezing and melting temperatures of liquids are the same	166.47	4	98.33	1	1.316	4	138
Heat energy is produced by work in overcoming friction	157.02	5	93.33	5	1.230	6	12
Condensation, a change from the vapor to the liquid or solid state, is caused by a decrease in the speed of the molecules	156.49	6	91.66	7	1.252	5	62
When a body is heated the average speed of molecular agitation increases	156.45	7	94.16	4	1.203	7	10

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Every pure liquid has its own specific boiling and freezing point	154.60	8	96.66	2	1.119	8	121
Heat is transferred by convection in currents of gases or liquids	145.77	9	90.83	8	1.061	9	72
Radiant heat travels in waves in straight lines; its intensity is inversely proportional to the distance to its source	144.75	10	92.50	6	1.009	10	89
Solids are liquefied and liquids are made into gases by heat; the amount of heat used in the process is specific but equals that given off in the reverse process	138.35	11	88.33	10	.966	11	68
The boiling of any liquid will take place at that temperature at which the pressure of the saturated vapor is equal to the pressure on the surface of the liquid	133.17	12	87.50	11	.882	13	123
Heat energy is produced by the compression of gases	130.57	13	86.66	12	.848	15	34
Heat energy is converted into mechanical energy by the expansion of gas	129.47	14	83.33	14	.891	12	64
Heat is conducted by the transfer of kinetic energy from molecule to molecule	122.81	15	77.50	18	.875	14	29
The total change in length of a bar is equal to its coefficient of linear expansion times the original length times the change of temperature in degrees C	118.58	16	80.00	16	.745	16	88
Diffusion and evaporation are the direct result of the high speed of molecular motion. They vary with temperature, area of exposed surface, and saturation and circulation of the gas in contact with evaporating liquid	110.45	17	80.83	15	.572	17	5
The weight of water per cubic meter of air needed to cause saturation depends on the temperature, being greater for higher temperatures	99.88	18	85.83	13	.252	19	140
If the volume of a confined body of gas is kept constant the pressure is proportional to the Kelvin or absolute temperature.	94.49	19	78.33	17	.312	18	86

Principle number	Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
	Light							
121	In a plane mirror the image lies as far behind the mirror as the object is in front of the mirror	183.75	1	95.83	4	1.698	1	112
72	An object appears to be a certain color because it reflects the wave lengths that produce that sensation and absorbs those which produce other color sensations	180.13	2	95.00	5	1.644	2	73
89	Light travels in straight lines in a medium of uniform optical density	179.89	3	99.16	1	1.559	4	111
68	A lens which is thicker in the center than at the edges converges a beam of light	175.51	4	94.16	6	1.571	3	115
123	A prism causes dispersion because the speed of light in glass is different for different colors of light	167.26	5	91.66	8	1.460	5	47
34	A lens which is thicker at the edges than in the center diverges a beam of parallel light	166.44	6	93.33	7	1.412	6	128
64	When a source of light is large in relation to the size of the object casting the shadow, the shadow will consist of an umbra and penumbra	162.99	7	96.66	3	1.281	9	141
29	The law of regular reflection states that the angle of incidence is equal to the angle of reflection	156.33	8	99.16	1	1.104	14	18
88	In a plane mirror, a line running from any point on the object to the image of that point is perpendicular to the mirror	155.02	9	89.16	11	1.272	11	126
5	When white light passes through a substance that absorbs some waves and not others, certain bands of color are missing with the production of an absorption spectrum	152.68	10	86.66	12	1.275	10	116
140	A concave mirror will produce a real, inverted image of an object when the object lies outside the principal focus	151.92	11	80.00	16	1.389	7	114
86	The quality of artificial illumination depends to a large extent on the degree to which the light has been diffused	146.74	12	97.50	2	.951	18	71
	A concave mirror will produce an erect, magnified, virtual image when the object is inside the principal focus	144.08	13	76.66	19	1.302	8	127

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Light from a point source is reflected from a concave mirror as a parallel beam when the source is at the principal focus	142.56	14	90.83	9	.999	16	33
When a source of light is nearer than the principal focus of a concave mirror the reflected beam is divergent	137.84	15	90.00	10	.924	20	42
When light enters a medium in which its speed is decreased, it is bent toward the perpendicular to the surface; when its speed is increased it is bent away from the perpendicular	136.06	16	77.50	18	1.131	13	44
When an object which is outside the principal focus is viewed through a converging lens a real, inverted image is formed on the opposite side of the lens	135.47	17	74.16	20	1.184	12	30
Refraction is caused by the change in speed of light as it leaves one medium and enters another	134.92	18	85.00	14	.964	17	20
Incandescent solids and liquids emit all kinds of light and give a continuous spectrum	132.73	19	84.16	15	.938	19	79
The illumination of a surface varies inversely as the square of the distance from the source.	132.17	20	85.83	13	.895	23	59
Luminous vapors and gases emit only certain kinds of light producing bright-line spectra.	125.92	21	79.16	17	.903	22	99
A diverging lens produces only erect, virtual images diminished in size	123.80	22	70.00	21	1.039	15	16
When a source of light is further away than the principal focus of a concave mirror the reflected beam is convergent	120.77	23	84.16	15	.707	27	19
When an object which is inside the principal focus and viewed through a converging lens the image is erect, magnified, and virtual	116.95	24	69.16	22	.923	21	36
When light is attempting to pass into a medium in which it will travel faster and when the angle of incidence is greater than the critical angle, it is totally reflected	111.29	25	65.00	24	.894	24	98
When white light passes through an incandescent gas or vapor the absorbed regions of the spectrum are narrow and a dark-line spectrum, Fraunhofer lines, is formed	108.35	26	67.50	23	.789	25	129

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
The index of refraction of any substance is the ratio of the speed of light in air to its speed in that substance	101.13	27	64.16	25	.714	26	76
Chromatic aberration is caused by the dispersion of white light when it passes through a lens	92.89	28	62.50	26	.587	28	78
Spherical aberration is caused by the failure of the marginal rays, passing through a lens, to pass through the principal focus	82.47	29	60.00	27	.434	29	45
Sound							
Musical sounds are distinguished from noises in that they have a definite pitch and are pleasing to the ear	178.48	1	97.50	1	1.564	2	15
The higher the pitch of a note, the more rapid the vibrations of the producing body	171.19	2	81.66	3	1.729	1	69
The speed of sound increases with an increase in temperature of the medium conducting it.	161.76	3	89.16	2	1.402	3	92
The loudness of a sound depends upon the amount of energy in the sound waves and decreases inversely as the square of the distance from the source	142.06	4	76.66	5	1.263	4	54
The speed of a wave is equal to the product of its frequency and wave length	111.87	5	75.83	6	.696	5	95
Two sound waves of the same or nearly the same frequency will destructively interfere with each other when the condensations of the one coincide with the rarefactions of the other	111.43	6	80.83	4	.591	7	75
When energy is transmitted in waves the medium which transmits the wave motion does not move along with the wave, but the energy does	104.40	7	73.33	7	.600	6	40
The frequency of vibration of a string is inversely proportional to its length, diameter, and square root of its density; and directly to the square root of the stretching force	75.62	8	65.83	8	.189	8	41

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Electricity							
Unlike poles of magnets attract each other; like poles repel each other	199.99	1	100.00	1	1.931	1	35
A magnet has two poles, connected by magnetic lines of force	188.49	2	100.00	1	1.709	3	13
A piece of iron or steel brought near the end of a magnet becomes magnetized by induction; the end of the iron near the pole of the magnet itself becomes a pole but unlike that of the magnet	180.87	3	90.83	9	1.739	2	63
A current-carrying coil of wire behaves like a magnet	164.53	4	94.16	7	1.359	4	93
A lack of electrons gives a body a positive charge	159.64	5	99.33	3	1.184	6	109
Magnets depend for their properties upon the arrangement of the molecules of which they are made up	158.71	6	90.00	10	1.327	5	66
In order to have a continuous current a closed circuit is necessary	153.05	7	96.66	4	1.089	7	56
All matter is made up of protons and electrons.	151.78	8	100.00	1	1.000	14	14
Charges on a conductor tend to stay on the surface and to be greatest on the sharp edges and points	148.75	9	95.83	5	1.022	11	108
There is always a loss in the form of heat when electrical energy flows through a conductor.	147.44	10	94.16	7	1.029	10	94
Electrical power in watts is equal to the amperes times the volts	143.79	11	89.16	11	1.055	8	17
Protons are found only in the nucleus of an atom	142.50	12	99.16	2	.837	19	91
Iron, cobalt, and nickel placed inside a coil carrying a current become magnetized by induction	142.46	13	95.00	6	.917	16	38
An electric current will flow when two metals of unlike chemical activity are acted upon by a conducting solution, the more active metal being charged negatively	141.20	14	88.33	12	1.021	12	80
In a series circuit all the current must flow through every part of the circuit	138.36	15	90.83	9	.918	15	97

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
In an uncharged body there are as many protons as electrons and the charges neutralize each other	134.92	16	95.00	6	.771	21	65
When a high-speed stream of electrons strikes a body, the atoms of that body are so agitated that they emit X-rays	128.39	17	75.00	20	1.031	9	143
Energy in kilowatt-hours is equal to the amperes times the volts times the time in hours, the whole divided by one thousand	125.99	18	74.16	21	1.001	13	74
In a parallel circuit there are a number of paths through which the current may flow; the current will divide, part going through each path	123.71	19	91.66	8	.619	26	4
When a current flows through a conductor magnetic lines are induced around the conductor, their direction depending upon the direction of the current	122.52	20	76.66	18	.847	18	27
The magnitude of the e.m.f. induced in a generator depends upon the number of turns of wire on the armature, the strength of the magnetic field, and the speed at which the armature turns	118.43	21	73.33	22	.871	17	132
The current induced in an armature is alternating but by means of a commutator it may be periodically reversed from the different coils to produce a direct current in the external circuit	114.80	22	77.50	17	.759	22	125
An e.m.f. is induced in a circuit whenever the number of magnetic lines of force cutting the circuit is changed	112.01	23	82.50	13	.570	29	21
When a current-carrying wire is placed in a magnetic field, there is a force acting on the wire tending to push at a right angle to the direction of the lines of force between the magnetic poles	110.95	24	69.16	25	.807	20	70
An electromagnet or coil exerts a choking effect on an alternating current due to the induced e.m.f.'s, produced by the changing field, tending to flow in the opposite direction	108.84	25	71.66	24	.718	23	67

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
In a series circuit the current is the same in all parts, the resistance of the whole is the sum of the resistances of the parts, and the voltage loss of the whole is the sum of the voltage losses of the parts	107.47	26	75.83	19	.611	27	53
Positively charged ions of metals may be deposited from an electrolyte to the cathode, when a current is sent through the solution	104.32	27	77.50	17	.518	32	139
A current-carrying coil of wire may be revolved and still not reverse its poles if the current is reversed periodically by a commutator	101.59	28	71.66	24	.578	28	110
The resistance of a conductor depends on the kind of material from which it is made, varies inversely with the cross-sectional area, and increases as the temperature increases	101.56	29	75.00	20	.513	33	119
The strength of an electric current varies directly as the e.m.f. and inversely as the resistance	98.63	30	79.16	15	.376	39	105
An insulated conductor may be charged by induction by grounding it and bringing a charged body close and then removing the ground. It will be charged oppositely to the charged body	98.49	31	72.50	23	.502	35	37
Alternating current charges a condenser twice during each cycle inducing an opposite charge on the other plate with the result that it appears that the current flows through the condenser	97.41	32	72.50	36	.481	23	43
Molecules lose electrons when struck by high-speed electrons or ions	97.18	33	61.66	33	.686	24	100
Condenser capacity varies directly with the area of the plates	95.84	34	80.00	42	.306	14	118
Gases at atmospheric pressure are poor conductors of electricity; they become conducting only when ionized	95.51	35	60.00	25	.684	35	48
Electrons are emitted from negatively charged heated terminals	94.78	36	66.66	31	.543	28	130

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
In a parallel circuit the total current is the sum of the separate currents, the voltage loss is the same for each branch, and the total resistance is less than the resistance of any one branch	93.42	37	74.16	40	.372	21	77
Electrostatic induction is the separation of charges on a conductor through the influence of a neighboring charge	92.47	38	89.16	50	.064	11	9
The amount of electrical energy in joules is equal to the power in watts times the time in seconds	90.86	39	78.33	45	.242	49	49
The magnitude of an induced e.m.f. is proportional to the rate at which the number of lines of force change and to the number of turns of wire that are cut by the lines	89.61	40	67.50	37	.427	27	46
In a tube which contains gas at a low pressure, cathode rays, streams of electrons, move away from the negatively charged terminal at high speed	89.305	41	68.33	38	.405	26	117
An induced current always has such a direction that its magnetic field tends to oppose the motion by which the current was produced	86.93	42	60.83	34	.504	34	96
The ratio of the impressed voltage of the primary to the induced voltage of the secondary of a transformer is equal to the ratio of the number of turns of wire in the primary to the number of turns of wire in the secondary	86.85	43	74.16	44	.245	21	24
A number of substances will emit electrons when negatively charged and when illuminated by light	84.83	44	56.66	30	.544	36	31
Gas at a low pressure is more easily ionized than is gas at ordinary pressure	77.62	45	63.33	43	.276	31	22
Heat energy in calories developed by a current is 0.24 times the current in amperes squared times the resistance times the time in seconds	74.46	46	62.50	46	.231	32	113

Principle	Combined index	Combined rank	Index of difficulty	Rank	Index of "Pleasure Value"	Rank	Principle number
Electrons of atoms are at a relatively great distance from the protons of the same atoms and are lost at times; they may also be added from other atoms	72.87	47	65.83	47	.136	29	124
Whenever a high-frequency oscillating current produces in the field around it oscillating electric and magnetic fields, radio waves are produced	69.94	48	51.66	41	.353	37	101
When an e.m.f. which has the same frequency as the natural frequency of a circuit is applied to that circuit, electrical resonance is produced	67.84	49	64.16	49	.071	30	81
By means of high-frequency generators or vacuum-tube oscillators, sustained or continuous, oscillations can be produced in a condenser circuit. Their intensity is made to vary with talking currents in a transmitter circuit to produce radio waves	54.97	50	50.00	48	.096	38	50

HABITS OF EYE-MOVEMENT IN READING*

E. DONALD SISSON

Wesleyan University

INTRODUCTION

In 1906 Dearborn[3] published the results of his pioneer study of eye-movements in reading, using the photographic recording technique devised by Dodge.[6] One of the most frequently quoted sections of this monograph deals with the concept of habit of eye-movement. Eye-movement habit, according to Dearborn, is evidenced by (1) "a rhythmical series of the same number of pauses per line" and (2) by a particular pattern of pause-fixations characterized by a relatively long initial pause followed by several of decreasing duration and terminated by a fairly long end pause. Although the text is by no means consistent on this point, the reader is led to believe that "short-lived motor habits," as Dearborn calls them, are in some way causally connected with good reading ability. In other words, the fast reader is said to be fast, in a large measure, because of his ability to set up habits of eye-movements more readily. It is also stated, in connection with a discussion of eye-movements and line length, that "short-lived motor habits" are favored by the shorter lines.

Many of the writers on the subject of reading, and especially of eye-movements in reading, subsequent to Dearborn have done lip service to his concept of eye-movement habit.† And, naturally enough, many of those interested in remedial reading work have attempted to improve reading performance via the eye-movement approach, in some cases endeavoring to instill regularity of fixations in line with the primary characteristic of the Dearborn "habit." Strangely enough, there has appeared no quantitative determination of either of the two features of the eye-movement "habit."‡ Proponents of the concept, and oppo-

* The author gratefully acknowledges his indebtedness to Dr. M. A. Tinker, University of Minnesota, in whose laboratory and under whose guidance this work was done.

† A review of the literature on the subject of habits of eye-movements will be found in the author's Ph. D. thesis on file at the graduate office of the University of Minnesota.

‡ Shen[11,15] found the variability of pause frequencies per line (*i.e.*, mean variability of all distributions of pause frequency per line) for Chinese material in horizontal and vertical arrangements. This variability is less for vertical than for horizontal, a fact which Shen accepts as support for his contention that the vertical is the more efficient and economical arrangement. He claims, "It is generally

nents too, have restricted themselves to general statements, theoretical considerations or assumptions.

The study to be reported here, utilizing quantitative analyses, aims at establishing the existence or non-existence of "short-lived motor habits" of eye-movement in reading, the relation of such "habits," if present, to the length of text line, to the type of material read, and to the level of reading ability.

PROCEDURE

Sixty subjects participated in this study. They were selected from the elementary laboratory sections in psychology at the University of Minnesota. The majority of them were, therefore, sophomores with a few juniors and seniors. Pre-experimental testing with the Chapman-Cook Speed of Reading Test[2] (time limit 1.5 minutes) showed that they comprised a fairly representative sample of reading ability (mean = 21.40 paragraphs, SD = 3.99). Each subject read twelve paragraphs of seven to eleven lines before the eye-movement camera.* Six of these were of easy narrative material, (from Stevenson's *Kidnapped*) in line lengths of nine, fourteen, nineteen, twenty-seven, thirty-two, and forty picas. The remaining six, in the same line lengths, were of scientific prose material (from Judd's *Psychology*). Each paragraph was preceded by a short practice paragraph of the same line length and of content logically preceding that of the test paragraph. Thus each subject read twenty-four paragraphs before the camera, and records were obtained for twelve of these. Half of the subjects read the narrative passages before the scientific prose, the order being reversed for the other half. After each paragraph, including the practice paragraphs, the reader was asked a question concerning its content. This was not supposed to serve as a measure of comprehension, but only as a means of setting up comparable reading attitudes in all subjects, and as an aid in preventing skimming.

The photographic records were analyzed for the usual measures: (1) Average number of fixations per line; (2) average number of regressions per line; (3) average pause duration; and (4) perception time per line. Furthermore, a measure of "habit" was devised as follows: A score of one was given to each succession of two lines which had the

---

conceded that any performance at the maximal efficiency will have a minimum of fluctuation."

*The Minnesota eye-movement camera has been described by Tinker.[18] For this experiment, timing was reduced to twenty-fifths of a second.

same number of fixations or came within one fixation of this number. This procedure is justified by Dearborn's description of the short-lived motor habit. It must be admitted that the allowance of a variation of one fixation is arbitrary. Its adoption was allowed as a compromise between the too rigid demand of identical number, and the too lenient allowance of a greater degree of freedom which would have obscured much of the individual variation. In order to reduce the habit scores from different paragraphs to a comparable basis, each paragraph score was divided by one minus the number of lines in the paragraph. In this way, an individual who reads all lines of a paragraph with the same number of fixations is assigned a "habit" score of 1.00, etc. This score was called the Habit Index.

RESULTS

The various eye-movement measures for the group are shown in Table I for the easy narrative material, and in Table II for the paragraphs of scientific prose. Perception time and pause duration are given in twenty-fifths of a second.

The first fact to be noted in these tables is that according to the eye-movement measures the scientific prose is not always more difficult than the easy narrative. In the nine- and thirty-two-pica lines the perception time per line is actually longer for the narrative material. Perception time, as Tinker[16,19] has pointed out, is practically a pure measure of reading rate. It represents about ninety-four per cent of the total reading time. And since there is no clear vision during the remaining six per cent (eye-movement time),[3,5,6,8,7] perception time is

TABLE I.—AVERAGES AND STANDARD DEVIATIONS FOR THE VARIOUS EYE-MOVEMENT MEASURES. EASY NARRATIVE MATERIAL IN VARIOUS LINE LENGTHS

Line length, picas	Number of fixations		Number of regressions		Perception time		Pause duration		Habit index	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
9	4.31	.90	.52	.43	25.05	5.01	5.85	.71	.685	.208
14	5.42	1.31	.62	.42	31.08	7.73	5.76	.68	.628	.203
19	7.31	1.36	1.10	.62	41.14	10.78	5.65	.67	.582	.188
27	9.03	1.63	1.30	.68	49.59	10.57	5.48	.65	.528	.222
32	11.22	3.88	2.19	1.02	64.75	15.29	5.62	.65	.298	.220
40	15.19	3.38	3.12	1.24	85.37	19.51	5.63	.80	.369	.203

an exact measure of the time consumed in acquiring and assimilating the textual material. Therefore, more stress is laid on perception time as a measure of reading ability than on any of the other measures. If, then, difficulty of text is defined in terms of rate of assimilation, we are led to the conclusion that in the nine- and thirty-two-pica lines, the easy narrative was more difficult of comprehension than the scientific prose. This point will be alluded to later.

TABLE II.—AVERAGES AND STANDARD DEVIATIONS FOR THE VARIOUS EYE-MOVEMENT MEASURES. SCIENTIFIC PROSE MATERIAL OF VARIOUS LINE LENGTHS

Line length, picas	Number of fixations		Number of regressions		Perception time		Pause duration		Habit index	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
9	3.83	.72	.47	.33	23.49	2.39	6.08	.74	.709	.219
14	5.86	1.13	.65	.44	34.40	7.50	5.88	.67	.578	.194
19	7.63	1.57	1.18	.68	44.53	12.39	5.97	.94	.436	.202
27	11.28	2.49	2.21	.94	67.01	12.27	6.07	.84	.343	.177
32	10.73	2.16	1.75	.93	60.22	13.03	5.64	.66	.405	.245
40	15.03	3.16	2.82	1.75	87.07	19.54	5.80	.68	.314	.165

A. HABIT AND LINE LENGTH

Dearborn[3] stated that "short lived motor habits" are more easily set up in short lines than in lines of greater length. A cursory glance at Tables I and II will reveal that the habit index decreases markedly from the nine-pica lines through the forty-pica lines. There should be no doubt about the fact. More questionable are the conclusions which Dearborn draws from this fact; that is, that the shorter line is more advantageous because, "first, its length is better suited to the formation of short-lived motor habits, and secondly, it has an advantage that it can (perhaps on this very account) be read in less time."

Examination of Tables I and II reveals that the relation between line length and habit index is not as close as would be expected if habit index were directly dependent on line length. In both tables there is a reversal in the habit index trend. For instance, the habit index for the thirty-two pica scientific prose paragraph is higher than that for the twenty-seven pica paragraph of the same kind of material. We have already pointed out that this paragraph is also read with fewer fixations than the preceding paragraph of shorter line length. In practically

every case, habit index varies with fixation frequency. Wherever scientific prose is read with fewer fixations than easy narrative, the habit index of the former is higher. Wherever a paragraph is read with fewer fixations than a selection of shorter line length, the habit index is higher for the one with fewer fixations. Habit index, then, seems to be a function of fixation frequency rather than line length, which, of course, would be highly correlated under ordinary conditions. In other words, there does not seem to be any factor intrinsic to line length itself that makes for greater ease of "short-lived-motor-habit" formation, but rather that higher habit indices in the short lines result fortuitously from the fact that these lines are read with fewer fixations. Where the number of fixations is small it is obviously impossible to deviate by very many fixations from line to line. Under these circumstances, chance would favor more sequences of identical or nearly identical numbers of fixations, simply because there are more lines of the same (or nearly the same) number of fixations. For illustration we have plotted the distributions of fixations per line for all the nine-, nineteen-, and forty-pica lines of the easy narrative material. The means and standard deviations of these distributions are given in Table III. Due to the fact that a lower limit is set by the necessity of

TABLE III.—MEANS AND STANDARD DEVIATIONS OF FIXATION PAUSES PER LINE FOR NINE-, NINETEEN-, AND FORTY-PICA LINE LENGTHS (EASY NARRATIVE)

Line length	Nine picas	Nineteen picas	Forty picas
Mean	4.31	7.30	15.19
SD	1.51	2.35	4.44

making at least one fixation per line, the distributions for both nine- and nineteen-pica lines are positively skewed.

These data indicate that 68.34 per cent of all the nine-pica lines deviate from the average in number of fixations by only 1.51 fixations, (in both directions). The same percentage of the forty-pica lines deviate by 4.44 fixations. It is surely much easier, by pure chance, to obtain more successive pairs of lines having the same or nearly the same number of fixations (habit index) in the short lines than in the longer ones.

We may conclude, then, that the numerical homogeneity of fixation frequency from line to line—that is to say, habit—is greater in the short than in the long lines. This effect, however, is a result, rather than a cause, of the fewer number of fixations required on the short lines.

B. TYPE OF READING MATERIAL

Is there any tendency for habit index to be favored more by one type of reading material than by another? We can answer this question by a brief reference to Tables I and II. Habit index, again, appears to vary inversely with fixation frequency, regardless of the type of reading material. Habit index does not depend on the type of material read except in so far as some materials are read with fewer fixations than others.

C. LEVEL OF READING PERFORMANCE

Dearborn's most frequently quoted conclusion is that good reading is characterized by a greater tendency to the formation of short-lived motor habits than is true of poor reading. In our analysis of the data regarding this point, we have used the eye-movement measure of perception time as the best single measure of reading ability. Furthermore, the data have been treated separately by line lengths. Accordingly, those subjects who read a nine-pica paragraph with a perception time shorter than the average for that paragraph, make up the group of good readers, for that particular paragraph, to be compared with the remainder of the readers. Comparisons were found by the group-difference method.

This analysis was made for all the eye-movement measures, but the results for habit index only will be given in detail here. In general, those readers with better (below average) perception times are also better than average in fixation frequency, regression frequency, and pause duration. In all cases the mean differences are in the expected direction, and in most cases (twenty-five of thirty-six differences) they satisfy the statistical requirements of significance. In the case of pause duration, only four out of the twelve differences are three times their standard errors. This is in keeping with the findings of previous investigators, notably Tinker,[16] Walker,[20] Buswell,[1] and Miles and Segal,[10] that pause duration is fairly constant for different kinds of material and varies little from one reader to another.

As regards habit index, the condition is even worse. Table IV presents the means for both groups, the mean differences, and the $D/\sigma D$ ratios, for each line length and for both kinds of reading material. Here, although all but one of the differences are in the same direction, only three of the twelve are statistically significant. It must be admitted, then, that if habit index is measuring any trait of reading ability,

TABLE IV.—MEANS, SD'S AND MEAN DIFFERENCE RATIOS OF HABIT INDEX FOR GOOD (BETTER THAN AVERAGE) AND POOR (WORSE THAN AVERAGE) READERS. VARIOUS LINE LENGTHS AND TWO TYPES OF READING MATERIAL

	Line length					
	Nine picas	Fourteen picas	Nineteen picas	Twenty-seven picas	Thirty-two picas	Forty picas
Easy narrative.						
M's good	.75	.72	.64	.58	.37	.43
SD	.21	.17	.19	.25	.23	.20
M's poor	.62	.50	.50	.48	.25	.31
SD	.19	.18	.15	.27	.20	.18
Difference	.12	.22	.14	.10	.12	.13
SE difference	.06	.05	.05	.08	.06	.05
$D/\sigma D$	2.25	4.34	2.96	1.31	1.92	2.80
Scientific prose.						
M's good	.79	.67	.51	.40	.40	.35
SD	.18	.16	.24	.19	.25	.16
M's poor	.63	.49	.36	.28	.41	.29
SD	.17	.18	.13	.13	.23	.16
Difference	.16	.18	.15	.12	−.01	.06
SE difference	.05	.05	.05	.05	.07	.05
$D/\sigma D$	3.10	3.48	2.74	2.58	.19	1.25

it is measuring it poorly. The only conclusion warranted is that readers better than the average in terms of perception time tend to be more regular in their eye-movements (have higher habit indices), but that the differences are very small and extremely unstable. The most probable explanation of this is the same as that given in connection with line length; namely, that since good readers make fewer fixations, they must of necessity deviate less from line to line in fixation frequency.

Poor readers usually make many fixations on some lines, but not on all. Thus their habit indices are low. For instance, one subject read the nineteen-pica easy narrative paragraph with the following numbers of fixations per line: fifteen, twelve, ten, thirteen, nine, six, seven, ten, twelve, eight. This performance yields an average fixation frequency of 10.2 fixations per line and a habit index of .222. A typical performance of a good reader on the same paragraph is nine, ten, ten, nine, eight, nine, eight, six, seven, seven fixations per line, yielding an average fixation frequency of 8.3 and a habit index of .889. It is

obvious that the good reader reads all lines well, whereas the poor reader makes many more fixations on some. If, now, we should reason that eye-movement regularity is a causal factor in reading ability, we might attempt to improve the performance of the poor reader by getting him to make the same number of fixations per line. This is the Dearborn concept of habit, and is the line of reasoning followed by Pressey,[12,13] Robinson,[14] and other proponents of habit. But on what level is he to stabilize his fixation frequency? He might, for instance, make fifteen pauses on each line. This would give him a high habit index, but he would be a poorer reader than now. So he must be persuaded to stabilize the number of his fixations at a low level in order actually to improve his reading. In other words, if the individual in question made fewer fixations per line, which by definition would make him a good reader, his habit index would rise; if he made the same number of fixations per line, which by definition would afford him a high habit index, his reading ability would not necessarily have improved. Whether or not he did improve would depend on the level at which he stabilized his fixation frequency.

D. HABIT INDEX AS A TRAIT

If training in the formation of eye-movement habit is to be effective, there must be some transfer from the specific drill situation to other reading performances. That is, habit must be a trait or human ability in the same way that intelligence, reading ability, or introversion are traits. Dearborn[3] was perhaps conscious of this problem of constancy when he wrote, "Because of the instability or lack of regularity in these rhythmical series, I have named them . . . , the 'short-lived' motor habits of the eye in reading" (p. 30). Dearborn certainly considers, however, that the ability to set up short-lived motor habits is characteristic of the good reader. If it is true, then it should be expected that those individuals with high habit indices in one reading situation would evidence high habit in other reading situations also.

This point has been tested by analysing the data to discover if those readers who are above the average in habit index on one paragraph are likewise above the average on other paragraphs. Scientific prose material was employed for this analysis using the nineteen-pica line length as the standard of comparison. Specifically, then, we have attempted to discover if those readers who are characterized by high habit indices on the nineteen-pica lines are also thus characterized on other scientific prose material. For purposes of comparison the

same analysis has been made in the case of perception time. Tables V and VI give these data. In Table V are shown the data for perception

TABLE V.—MEAN DIFFERENCE RATIOS OF PERCEPTION TIME FOR VARIOUS PARAGRAPHS, FOR GOOD READERS AND POOR READERS ON THE NINETEEN-PICA PARAGRAPH. SCIENTIFIC PROSE MATERIAL

	Line length					
	Nine picas	Fourteen picas	Nineteen picas	Twenty-seven picas	Thirty two picas	Forty picas
M's of good readers....	20.38	30.76	37.17	60.40	52.56	74.10
SD.....	4.01	4.89	5.38	7.41	9.16	14.75
M's of poor readers....	23.34	36.85	52.20	70.64	66.52	90.74
SD.....	3.72	7.65	13.01	12.67	11.96	16.21
Difference.....	2.96	6.09	15.03	10.24	13.96	16.64
SE difference.....	1.22	2.06	2.93	3.34	3.65	5.76
$D/\sigma D$.....	2.43	2.96	5.13	3.06	3.82	2.89

time. In every case the differences are in the expected direction, and all of them are fairly high, those close to the standard (nineteen picas) being more significant than those farther away.

When we turn to Table VI, however, we find a different situation. The differences are in both directions, and (aside from the case of the nineteen-pica line where an exception is expected) none of the $D/\sigma D$

TABLE VI.—MEAN DIFFERENCE RATIOS OF HABIT INDEX FOR VARIOUS PARAGRAPHS, FOR READERS WHO ARE ABOVE AND BELOW AVERAGE HABIT INDEX ON THE NINETEEN-PICA PARAGRAPH. SCIENTIFIC PROSE MATERIAL

	Line length					
	Nine picas	Fourteen picas	Nineteen picas	Twenty-seven picas	Thirty-two picas	Forty picas
M's of good readers....	.759	.641	.643	.285	.369	.257
SD.....	.203	.229	.154	.168	.229	.200
M's of poor readers....	.682	.572	.298	.368	.389	.391
SD.....	.159	.170	.111	.166	.237	.123
Difference.....	.077	.069	.345	−.083	−.020	−.134
SE difference.....	.062	.076	.042	.056	.080	.060
$D/\sigma D$.....	1.240	.910	8.210	1.480	.250	2.230

ratios is high enough to warrant the attachment of statistical significance. The two greatest differences are not in the expected direction. In one case (forty pica), the chances are ninety-nine in one hundred that the true difference is greater than zero. That would mean that those readers who are characterized by most regularity and rhythmicity on the nineteen-pica lines are least rhythmic on the forty-pica lines. We cannot say, therefore, that habit index measures any trait or ability which individuals possess in varying degrees. Rather it appears that habit index measures the concatenation of chance events which causes sequences of same-number-of-fixation lines in a given paragraph. We have already indicated how such an explanation could account for the relation, in any specific situation, between habit index and level of reading ability, and also habit index and line length. Some such explanation seems necessary to fit the facts. If habit index were related to level of reading ability in some truly etiological manner (in the sense that one was a necessary cause of the other), then we might expect to be able to predict habit index on the basis of performance on nineteen-pica lines, as is possible in the case of perception time. The fact that there is no relationship which permits of prediction from one situation to another in respect to habit index, indicates that this measure is not a habit trait, and is only accidentally related to reading ability.

E. TEMPORAL DISTRIBUTION OF PAUSES

The second characteristic of "short-lived motor-habits," one which has not been emphasized by either Dearborn or subsequent investigators, refers to the temporal distribution of fixation pauses within the line. The two aspects of this temporal arrangement are, according to Dearborn,[3] (p. 108), "(1) the first pause of a line longer than succeeding pauses; and (2) a secondary increase in the duration of pauses near the end of the line." The long initial pause permits a kind of general survey of the line prior to actual reading. The secondary increase in pause duration toward the end of the line serves, according to Huey,[9] perhaps for review, although the reasons for this suggestion are not clear.

A check for this factor of temporal distribution has been attempted through the following analysis. The five best and the five poorest readers were chosen on the basis of perception time for each of the paragraphs of easy narrative. For each of these groups were found the average duration of the first, second, next to last, and last pauses in

the line. These data are given in Table VII. In the case of the extremely short nine-pica lines, only the first, second, and last fixations could be averaged, since many of the good readers made only three fixations on these lines.

In general, these data show the initial fixations to be longer than the others, but there is no consistent tendency for the last fixation to be longer than the others. In the majority of cases, the last fixation is, in fact, shorter than all of the others. In the longer lines (thirty-two and forty picas) the second fixation is quite markedly longer than the first. This is quite evidently explained by the fact that most of the first fixations of these long lines are premature stops in the return sweep, and so must be followed by a regression to the proper place at the beginning of the line. The first fixation, then, would be only long enough

TABLE VII.—AVERAGE DURATION (IN TWENTY-FIFTHS OF A SECOND) OF FIRST, SECOND, NEXT TO LAST, AND LAST PAUSES, FOR DIFFERENT LINE LENGTHS, AND FOR GOOD AND POOR READERS. ($N = 5$)

	Fixations			
	First	Second	Next to last	Last
Nine pica.				
Good	6.0	5.5	...	5.4
Poor	6.5	5.9	...	6.8
Fourteen pica.				
Good	6.4	4.9	5.0	4.8
Poor	7.6	5.9	6.5	5.8
Nineteen pica.				
Good	6.8	5.5	4.9	4.7
Poor	6.5	5.7	5.7	5.6
Twenty-seven pica.				
Good	5.9	5.0	4.6	4.3
Poor	6.2	6.7	5.1	4.6
Thirty-two pica.				
Good	4.9	5.7	5.2	5.5
Poor	6.2	6.6	5.8	5.3
Forty pica.				
Good	5.1	6.0	4.5	4.1
Poor	4.1	7.6	5.3	5.5

for the reader to become aware of the fact that a further return in the same direction was necessary. The true initial reading fixation, in such instances, is the second fixation.

Aside from this unimportant detail, our data would show no difference between the temporal pattern of the short lines as compared to the longer ones. In all cases the fixation nearest the beginning of the line (*i.e.*, initial reading fixation) is consistently of longer duration than the others; also, there is a marked tendency for the last fixation to be shorter than the average. These data contradict Dearborn's statement that the particular mode of time distribution, which is part of his habit concept, is set up more easily in the short lines. Practically no differences in temporal pattern occur from nine through forty picas with the exception mentioned above. Since such is the case, Dearborn's suggested explanation must be discarded. He assumed that since the feature of long initial pause was found only, or oftener, in the short lines, it might serve the purpose of a preliminary survey by means of peripheral vision of the whole line before reading it—a performance which, obviously, would be impossible in the very long lines. Further contradictory evidence is that there seems to be no relation between the duration of the initial pause and the length of the return sweep. According to the data presented in Table VII the initial pause is as long in the short lines as it is when a return sweep of thirty-two or forty picas is necessitated. Moreover, it is of about the same duration after the very short refixations necessitated by the long lines.

Our data also contradict the claim of Dearborn that this mode of temporal distribution is related to the level of reading ability. According to Table VII there is no significant difference in this respect between the best and the poorest readers among our subjects. The differences between the groups in reading ability are extreme, yet the differences in regard to pattern of pause distribution are non-existent or negligible. And if this pattern is the same regardless of line length, and is as typical of poor readers as of good, it is difficult to see how it can bear any causal relation to the regularity feature of habit which varies with both of these factors. It is unjustifiable, then, to make temporal distribution of pauses a factor of causal significance in reading ability.

F. SUMMARY

The concept of habit of eye-movements, as revealed in the literature, posits a regularity of fixation frequency from line to line, and a particular mode of temporal distribution of pauses within the line. The formation of this eye-movement habit is reported to be favored by short lines; moreover, those individuals whose reading performance

is characterized by more of this habit factor, are reported to be able to read more rapidly on this very account.

Our aim in this experiment has been to make an experimental analysis of the habit concept in order to discover if these things are true. Eye-movement records were obtained for the reading of paragraphs of easy narrative material and of scientific prose in six-line lengths. A measure of habit was derived from the descriptions of the concept found in the literature. This measure is called the habit index. The main facts brought out by our analysis are listed below.

1. Habit index, defined as the percentage of successive pairs of lines having the same or nearly the same number of fixations, is higher for the short lines than for the long ones.

2. This fact is accounted for by probability alone without necessitating the assumption of some factor implicit in short lines which makes for greater ease of habit formation.

3. There is no consistent tendency for habit index to favor one kind of reading material more than another.

4. Good readers tend to have a higher habit index than poor readers, although the differences are not, by and large, statistically significant. The reason for this relation is that habit index is a function of the number of fixations made. That is, the fewer the fixations made, the better the chance of two adjacent lines having the same number of fixations.

5. Habit index cannot be said to measure any individual trait, since there is little, if any, consistency from one situation to another. In some cases, those who have high habit indices on the standard (nineteen pica) line have indices lower than average on other lines. This measure, then, is not a habit trait.

6. In general, initial reading pauses are longer than average, but final pauses appear to be briefer than average.

7. Short lines do not differ from long lines in regard to temporal distribution of fixation pauses within the line, nor do good readers differ from poor readers in this respect.

8. Pattern of temporal distribution does not seem to be related in any way to habit index, nor does it appear to have any value as a part of an eye-movement habit concept.

9. In general, a high habit index seems to be a consequent of a low fixation frequency, and a result, rather than a cause, of rapid reading. In any event, the concept of "short-lived motor habit" appears to be a useless one, since neither of its characteristics (high line-to-line regular-

ity of reading pauses and particular pattern of temporal distribution of these pauses within the line) is a consistent differentiating factor.

BIBLIOGRAPHY

1. Buswell, G. T.: "Fundamental reading habits: a study of their development." *Suppl. Educ. Monog.*, 1922, No. 21.
2. *Chapman-Cook Speed of Reading Test.* Philadelphia: J. B. Lippincott Co.
3. Dearborn, W. F.: "The Psychology of reading." *Columbia Univ. Contr. to Philos. and Psychol.*, Vol. XIV, 1906, No. 1.
4. Dodge, R.: "Visual perception during eye-movement." *Psychol. Rev.*, Vol. VII, 1900, pp. 454–465.
5. Dodge, R.: "The illusion of clear vision during eye-movement." *Psychol. Bull.*, Vol. II, 1905, pp. 193–199.
6. Dodge, R. and Cline, T. S.: "The angular velocity of eye-movements." *Psychol. Rev.*, Vol. VIII, 1901, pp. 145–157.
7. Holt, E. B.: "Eye-movement and central anaesthesia." *Psychol. Rev., Monog. Suppl.*, Vol. IV, 1903, pp. 3–45.
8. Holt, E. B.: "Visual sensation and eye-movement." *Psychol. Bull.*, Vol. II, 1905, pp. 388–390.
9. Huey, E. B.: *The Psychology and Pedogogy of Reading.* New York: The Macmillan Co., 1908.
10. Miles, W. R. and Segal, D.: "Clinical observations of eye-movements in the rating of reading ability." *J. Educ. Psychol.*, Vol. XX, 1929, pp. 520–529.
11. Miles, W. R. and Shen, E.: "Photographic recording of eye-movements in the reading of Chinese in vertical and horizontal axes: method and preliminary results." *J. Exper. Psychol.*, Vol. VIII, 1925, pp. 344–362.
12. Pressey, L. C.: *A Manual of Reading Exercises for Freshmen.* Columbus: 1928.
13. Pressey, L. C. and Pressey, S. L.: "Training college freshmen to read." *J. Educ. Res.*, Vol. XXI, 1930, pp. 203–211.
14. Robinson, F. P.: "The rôle of eye-movements in reading with an evaluation of techniques for their improvement." *Univ. of Iowa Stud.*, 1933, No. 39.
15. Shen, Eugene: "An analysis of eye-movements in the reading of Chinese." *J. Exper. Psychol.*, Vol. X, 1927, pp. 158–183.
16. Tinker, M. A.: "Eye-movement duration, pause duration, and reading time." *Psychol. Rev.*, Vol. XXXV, 1928, pp. 385–397.
17. Tinker, M. A.: "A photographic study of eye-movements in reading formulae." *Genet. Psychol. Monog.*, Vol. III, 1928, No. 2.
18. Tinker, M. A.: "Apparatus for recording eye-movements." *Amer. J. Psychol.*, Vol. XLIII, 1931, pp. 116–118.
19. Tinker, M. A. and Frandsen, Arden: "Evaluation of photographic measures of reading ability." *J. Educ. Psychol.*, Vol. XXV, 1934, pp. 96–100.
20. Walker, R. Y.: "The eye-movements of good readers." *Psychol. Rev.*, Vol. XLIV, 1933, pp. 95–117.

RETROACTIVE INHIBITION IN RELATION TO AGE AND THE NATURE OF THE INTERPOLATED TASK

T. G. FORAN

The Catholic University of America

The phenomenon of retroactive inhibition has been the subject of numerous investigations, particularly within recent years. Britt's comprehensive review of the literature[1] includes a bibliography of one hundred forty-one references, some of which are more specifically concerned with other aspects of forgetting. Since Britt's review was published in 1935, additional investigations have shed further light on the rôle of some of the many variables involved. Few studies, however, have dealt with retroactive inhibition in the learning of children, and in nearly all of the studies the records of the children's performance were grouped with those of the adults. The only study directly related to the present investigation is that of Dreis[2] who concluded that susceptibility to retroactive inhibition bore no relation to age.

There are no apparent and well-established reasons why the results obtained from experiments with adults should or should not apply to the learning of children. It is theoretically possible that mental adaptations can be made either more or less readily by children and arguments of one kind or another might be advanced in support of either contention.

The experiments conducted with adults have dealt with many phases of the general problem, especially with the relation between the amount of retroactive inhibition and the similarity between the original and interpolated activities. Similarity has generally been recognized as one of the major conditions of retroactive inhibition. Robinson has formulated the widely accepted principle that as similarity between interpolation and original memorization is reduced from near identity, retention falls away to a minimum and then rises again, but, with decreasing similarity, it never reaches the level obtaining with maximum similarity.[3] Skaggs has defined the relation between similarity and retroactive inhibition in five principles[4] that are in essential agreement with Robinson's law. Some experimental results have not confirmed the general principle but the crux of the problem is what constitutes similarity. Robinson and others have recognized the fact that in such complex activities there are many elements and conditions precluding the identification of similarity with either the content or the function alone or together. Some of

the conflicting evidence might be interpreted to mean that other factors are more important than similarity or that similarity is too vague to describe adequately the relations between two learning activities.

The problem in the present investigation was to determine the extent of retroactive inhibition induced by various kinds of interpolated activities at ages eight to thirteen, inclusively.* The original learning consisted in the memorization of twenty-five verbs each containing from five to seven letters, inclusively. The words used were selected from the first two thousand five hundred words in Thorndike's *Teacher's Word Book.* Familiar words were chosen to avoid the distractions and errors arising from the use of unknown words.

The control group engaged in physical exercises during the fifteen minute interval between the first and the delayed recall tests. This form of activity would seem to furnish a closer approximation to rest than would any other than could be arranged under the classroom conditions that prevailed. There is, of course, no guarantee that any activity will exclude completely all reference to the previous work, but the consistency of the results obtained and the type of activity engaged in warrant the assumption that the interval did not provide opportunity for other than the normal amount of forgetting. The children had, of course, no intimation that a second test would be given.

The pupils in grades III to VIII, inclusively, in one school formed the control groups for each of the age levels from eight to thirteen. Six other schools formed the experimental groups in which some form of mental activity required by the assignments filled the interval between the first and delayed recall tests. The six forms of interpolated activity were arithmetic computation, arithmetic problem-solving, spelling, memorizing a poem, memorizing a list of nouns, and, in one school, computation and spelling, half the interpolated time being devoted to each. In this school the procedure differed notably from the usual arrangement of conditions in experiments of this type, since the inclusion of two exercises involved the same amount of interpolated time but additional adaptations to the tasks. This arrangement was prompted by the supposition that the amount of retroactive inhibition would be increased if the phenomenon in question is due to interference with perseveration rather than to transfer.

* This is the first of a series of studies to be reported on retroactive inhibition in the learning of children. Succeeding reports will deal with the influence of such other variables as the degree of difficulty of the interpolated activity, the number of interpolated tasks, etc.

The scores of pupils who were less than eight or more than fifteen years of age were discarded, as such pupils are not representative of children of those ages. The control group did not include a sufficient number of fourteen-year-old pupils to permit measures of retroaction at that age, but the results from the experimental groups are reported for that age level, the final measure of retroactive inhibition being omitted. The total number of pupils in this experiment is one thousand seven hundred thirty-three. There is some reason to question the dependability of the results at the eight-year level, for pupils of that age in the third grade in March are above the average in intelligence. A study, as yet unpublished, by one of the writer's students indicates that intelligence is a factor in susceptibility to retroactive inhibition.

The experiments were conducted as group studies. The instructions were comparatively simple. Each pupil was furnished with a booklet containing the list of words to be learned, a blank page for the first test, pages containing the material for the interpolated activity, and another blank sheet for the final test. Five minutes were allowed for the original learning and four minutes for each of the two tests. The interpolated activity consumed fifteen minutes, including time for the directions and all other work connected with the intervening task. The time limits were suited to the difficulty of the original task for there were no zero scores and only a few perfect scores.

The scoring of the tests ignored the spelling of the words as long as there were reasonable indications that the word in question was one of the original list. The use of derived forms was permitted unless there was considerable difference between the words. As the scoring was in accordance with definite rules and all done by persons who had direct experience with such work, the subjective element has been reduced to a minimum.

Table I contains the original and delayed recall scores of the children in the control groups. In all the experimental groups, the measure of retroactive inhibition is that commonly used: the difference—percentage of recall of the control group minus the percentage of recall of the experimental group—divided by the percentage of recall of the control group.

$$\text{per cent RI} = \frac{\text{per cent } R_c - \text{per cent } R_e}{\text{per cent } R_c}$$

This measure of retroactive inhibition allows for differences in the means of the original scores of control and experimental groups and,

therefore, dispenses with the need of equating such groups. Other studies, as yet unreported, reveal that equating is superfluous when using this measure of retroactive inhibition and having groups as large as those used in this series of experiments.

Arithmetic Problem-solving.—The results obtained from the groups who spent the intervening time solving arithmetic problems are summarized in Table II. No significant differences have been obtained, although some of the measures show slight degrees of retroactive inhibition or of facilitation. On the whole, however, the results indicate rather definitely that the solving of arithmetic problems neither aids

TABLE I.—INITIAL AND DELAYED RECALL OF CONTROL GROUPS

	Chronological ages					
	8	9	10	11	12	13
N	23	42	38	39	43	41
M_1	5.44	8.26	9.61	10.19	10.73	12.35
	.51	.66	.75	.64	.63	.59
M_2	4.91	7.45	8.61	9.63	9.48	10.70
	.49	.56	.71	.65	.59	.63
M_2/M_1	90.3	90.2	89.6	94.5	88.4	86.6

TABLE II.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITY WAS ARITHMETIC PROBLEM-SOLVING

	Chronological ages						
	8	9	10	11	12	13	14
N	33	63	59	72	76	78	24
M_1	6.68	7.61	6.74	10.28	10.37	12.83	12.83
	.68	.57	.40	.44	.48	.46	.68
M_2	6.14	7.29	6.02	9.06	9.18	10.83	11.00
	.69	.57	.43	.40	.49	.46	.79
M_2/M_1	91.9	95.8	89.3	88.1	88.5	84.4	85.7
Per cent RI	+1.8	+6.2	0.3	6.8	+0.1	2.5	

nor hinders the recall of a series of words that had been memorized immediately previous.

Computation.—Table III contains the immediate and delayed recall scores of the children who performed arithmetic computations. At ages eight, nine, and ten there are appreciable degrees of facilitation and at ages eleven, twelve, and thirteen, small amounts of interference. There is no evident reason for the reversal of the trend. It will be

TABLE III.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITY WAS COMPUTATION

	Chronological ages						
	8	9	10	11	12	13	14
N	25	25	35	28	37	25	9
M_1	6.74	11.46	11.24	15.86	16.01	16.66	15.33
	.50	.76	.69	.69	.67	.79	1.30
M_2	6.34	12.66	11.13	14.14	14.12	14.10	12.33
	.59	.84	.74	.87	.72	.81	.85
M_2/M_1	94.1	110.5	99.0	89.2	88.2	84.6	80.4
Per cent RI	+4.2	+22.5	+10.5	5.6	0.2	2.3	

observed that the initial recall scores of the pupils in this school are unusually high, suggesting the possibility that these pupils were more accustomed than others to such tasks as that involved in the original learning. Skaggs has noted that practice with any material tends to reduce the influence of factors making for retroaction.[4] It is also known that the degree of initial learning modifies the extent of the retroactive inhibition. Either or both of these factors may be responsible for the results obtained in this school. The degrees of retroactive inhibition with other types of interpolated tasks do not show that there is much greater susceptibility to retroactive inhibition at some ages rather than others, and, as the effect is measured relatively, it would appear to be due to some such influence as that of habituation to the tasks that the experiment presented.

Spelling.—Table IV contains a summary of the results obtained with spelling as the interpolated task. Substantial amounts of retroactive inhibition are observed at the eight-, nine-, and ten-year levels

and increasing degrees of facilitation at eleven, twelve, and thirteen years. There is pronounced consistency in the decreasing degrees of inhibition culminating in equally significant degrees of facilitation with increasing age. The same spelling test was used in all classes. It was formed by combining two of the lists in the Morrison-McCall Scale. It is possible that the spelling task, being more difficult for the younger children, thereby caused interference at the earlier ages, whereas the words presented no special difficulty for the older children who were not only not distracted but actually assisted by the writing or other elements that the spelling test involved. The percentages of pupils obtaining the same, lower, or higher scores on the second test confirm the trends shown by the mean scores and the percentages of recall and retroactive inhibition.

TABLE IV.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITY WAS SPELLING

	Chronological ages						
	8	9	10	11	12	13	14
N	18	29	34	37	24	37	16
M_1	7.28	11.05	11.68	13.80	14.50	15.58	17.75
	.57	1.09	.80	.82	1.07	.75	1.42
M_2	5.06	8.16	9.85	13.69	14.08	15.91	17.38
	.60	1.03	.89	.77	1.36	.71	1.44
M_2/M_1	69.5	73.8	84.3	99.2	97.1	102.1	97.9
Per cent RI	23.0	18.2	5.9	+5.0	+9.8	+17.9	

	Chronological ages						
	8	9	10	11	12	13	14
Per cent same score	5.6	13.8	17.6	18.9	12.5	18.9	12.5
Per cent losing	88.9	75.9	67.6	48.6	45.8	37.8	50.0
Per cent gaining	5.6	10.3	14.7	32.4	41.7	43.2	37.5

These results indicate the desirability of investigating the influence of several factors that are in some way related to age, ability, and

maturity. Such factors appear to be significant elements in the production of retroactive inhibition, possibly through the difficulty of the interpolated task.

Memorizing Poetry.—Table V contains a summary of the results obtained when the interpolated activity was the memorization of a poem. There is a small amount of inhibition at the eight-year level and small and probably insignificant degrees of facilitation at ten, eleven, twelve, and thirteen. None of the measures of interference or facilitation is large enough to warrant attributing any effect to the interpolated activity. The absence of any interference is noteworthy, since the task consisted in the verbatim memorization of a poem. The results suggest that similarity, conceived in terms of the functions required by the two learning activities, is either not a condition of retroactive inhibition or is so concealed by other factors that its effect is obliterated. At the same time it may be mentioned that interference could hardly be compensated by other factors if it consists, as some hold, in the blocking of perseveration.

Memorizing Nouns.—The experimental situation that most closely resembles the original learning prevailed when the interpolated task

TABLE V.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITY WAS MEMORIZING A POEM

	Chronological ages						
	8	9	10	11	12	13	14
N	27	53	59	58	52	40	14
M_1	8.28	9.67	11.96	12.40	15.27	12.45	12.50
	.65	.54	.54	.72	.65	.72	.83
M_2	6.87	8.65	11.25	12.19	13.92	11.60	13.50
	.66	.56	.63	.77	.67	.74	1.19
M_2/M_1	83.0	89.5	94.1	98.3	91.2	93.2	108.0
Per cent RI	8.1	0.8	+5.0	+4.0	+3.2	+7.6	

was the memorization of a list of nouns. The results are shown in Table VI. Large degrees of retroactive inhibition occur at all age levels. No other interpolated activity induced as much interference as did this one. The amounts of inhibition do not show any consistent

relation to age but the exceptionally high initial mean score of the eight-year-old pupils introduces a probable source of error. The initial score of the nine-year-old pupils is also high, and the mean score of the ten-year old pupils is relatively low. The extent to which such irregularities in the original means affect the measures of retroactive inhibition cannot be appraised from the available data. The smallest amount of inhibition occurs in the scores of the thirteen-year old pupils and the percentages of retroactive inhibition for the ten-, eleven-, and twelve-year pupils are practically identical. If age is not associated with retroactive inhibition in this phase of the experiment, it would seem that the effects of other interpolated tasks would have to be traced to such conditions as the familiarity of the work, the degree of attention demanded, the difficulty of the task, etc.

Spelling and Computation.—As already stated, one school divided the fifteen-minute interval evenly between spelling and computation. Such an experimental situation was arranged to test the hypothesis that the number of interpolated tasks might be more important than their kind. Table VII presents the results. It will be seen that the measures of the net effect of the two tasks are small and that there is actually slight facilitation at the eight-year level. The results obtained do not warrant any conclusions and the hypothesis remains to be tested by further experimentation. It is likely that much depends on which of the tasks follows immediately the original learning and

TABLE VI.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITY WAS MEMORIZING A LIST OF NOUNS

	Chronological ages						
	8	9	10	11	12	13	14
N	22	34	35	21	38	45	21
M_1	10.41	9.50	8.90	9.74	11.71	13.48	14.60
	1.42	.68	.88	.55	.72	.59	.90
M_2	6.14	4.21	4.79	5.36	5.97	8.09	11.83
	1.11	.67	.80	.84	.65	.79	1.48
M_2/M_1	59.0	44.3	53.8	55.0	51.0	66.0	81.0
Per cent RI	34.7	50.9	40.0	41.8	42.3	23.8	

which immediately precedes the delayed recall test. There is no similarity whatever between the results when spelling and computation were both included and when each of them was used alone.

TABLE VII.—INITIAL AND DELAYED RECALL WHEN INTERPOLATED ACTIVITIES WERE SPELLING AND COMPUTATION

	Chronological age						
	8	9	10	11	12	13	14
N	16	48	28	43	35	21	13
M_1	7.25	9.46	10.36	11.99	11.99	13.64	14.35
	.71	.64	.89	.87	.68	.66	1.03
M_2	7.12	8.37	9.36	10.31	10.10	11.36	11.42
	.70	.57	.95	.78	.90	.98	1.16
M_2/M_1	98.2	88.5	90.3	86.0	84.2	83.3	79.6
Per cent RI	+8.7	1.9	+0.8	9.0	4.8	3.8	

The results of this series of experiments invite further study of the problems considered. It does appear from the data obtained that there are significant factors affecting the influence of one learning activity on another besides the objective resemblance of the two tasks. Among such conditions of retroactive inhibition appear to be the ease or difficulty of the interpolated task and the extent to which it is habituated. Such characteristics may be regarded as additional elements to be considered in describing the similarity between the original and interpolated tasks.

SUMMARY AND CONCLUSIONS

Comparisons at ages eight to thirteen, inclusively, between retention of control and experimental groups of a list of twenty-five words indicate varying effects produced by incompletely identified attributes of the interpolated activity.

(1) Arithmetic problem solving and memorizing poetry produced no significant effects on retention of the list of words learned previously. The tendency observed was towards slight degrees of facilitation at all except the lowest age level when poetry was the interpolated learning.

(2) Interpolated computation produced facilitation at eight, nine, and ten years and negligible degrees of retroactive inhibition at eleven, twelve, and thirteen years. The absence of retroactive inhibition in the learning of the younger children is possibly due to their marked proficiency in memorizing words as shown by the initial scores although the older children in the same experimental group had high initial scores also.

(3) At eight, nine, and ten years, spelling produced marked but decreasing degrees of interference, and at eleven, twelve, and thirteen, increasing amounts of facilitation were noted. The tentative explanation is that susceptibility to retroactive inhibition is conditioned by the habituation of the interpolated task.

(4) The maximum amount of inhibition occurred when the interpolated activity was very similar to the original, both involving the memorization of lists of words.

(5) In the maximum degrees of retroaction produced by the interpolated memorization of other words, no relation appears to exist between age and the amount of such interference.

(6) The inclusion of two activities produced no consistent results that could be related to those induced by such activities separately.

REFERENCES

1. Britt, S. H.: "Retroactive Inhibition: A Review of the Literature." *Psychological Bulletin*, Vol. XXXII, 1935, pp. 381–440.
2. Dreis, T. A.: "Two Studies in Retroaction: I. Influence of Partial Identity. II. Susceptibility to Retroaction at Various Grade Levels." *Jour. General Psychol.*, Vol. VIII, 1933, pp. 157–172.
3. Robinson, E. S.: "The Similarity Factor in Retroaction." *Amer. Jour. Psychol.*, Vol. XXXIX, 1927, pp. 297–312.
4. Skaggs, E. B.: *The Major Forms of Inhibition in Man.* Chicago: Univ. of Chicago Press, 1931. Pp. vii + 104.

IMPROVING ABILITY IN THE MATHEMATICAL CONCEPTS OF EDUCATIONAL STATISTICS

GEORGE E. HILL
University of Pennsylvania

It has long been known that many college students labor under the handicap of an inadequate command of the basic skills in arithmetic. A study[1] of college students has revealed that many exhibit distinct arithmetical disabilities, although it is doubtful that they all need systematic review of elementary arithmetic.

That students in courses in educational statistics also lack adequate command of the mathematical processes used in such courses has also been demonstrated. Brown[2] gave a test devised to measure skill in the mathematical processes called for in statistics texts to nearly one thousand students and found that their scores on the test correlated +.44 with their marks in statistics. He concluded that "mathematical competence is an important factor, although by no means the sole factor in determining success in the statistics course."

The writer was interested to determine what effect skill in the simple arithmetic and algebraic processes used might have upon success in elementary educational statistics when the students were helped to overcome initial handicaps. It was his experience that students entering such a course varied widely in their command of the necessary mathematical processes.

This report describes the writer's efforts to overcome these weaknesses with two classes in "Elementary Educational Statistics" totalling fifty-two students. The classes met once a week for two hours over a period of sixteen weeks. The students were all graduate students with diverse backgrounds, though almost all of them were teachers in service and school administrators. Their skill in the mathematics used in the course was determined by a test based on the arithmetic and algebraic processes found in the textbook which was used. This test consisted of twenty simple problems very similar to those found in the test used by Brown.[3] It should be noted that it was not a general mathematics test, but a test of the particular skills

[1] Arnold, H. J.: "Arithmetical Abilities and Disabilities of College Students." *Elementary School Jour.*, Vol. XXI, 1930, pp. 259–270.

[2] Brown, Ralph: *Mathematical Difficulties of Students of Educational Statistics.* Contributions to Education, No. 569, Teachers College, Columbia University, 1933.

[3] Brown, Ralph: *Op. cit.*, pp. 13–14.

used constantly in the course. The processes included are listed below with one example of the type of problem used for each:

1. Four fundamental processes with whole numbers.
   Example: $14 + 6(7) - 20 = ?$
2. Changing common fractions to decimal fractions and percentage.
   Example: 42 is what per cent of 126?
3. Four fundamental processes with decimal fractions.
   Example: $29.8 + 31.7 = ?$
4. Four fundamental processes with signed numbers.
   Example: $32 - (-2)(-4) = ?$
5. Square Root.
   Example: The square root of 125.44 is what?
6. Substitution in an algebraic formula.
   Example: If $a = 2$, $b = 7$, and $c = .5$

$$ab - \frac{cb}{a} = ?$$

This test was given at the first class meeting and again at the time of the final examination. The first administration is here referred to as Test I and the final administration as Test II. The same test was used both times.

The students' success in the statistics course was measured by their scores on three periodic examinations and their ability to work ten long problems which constituted the practice content of the course. The examination scores totaled two hundred fifty points for perfect performance, the problems two hundred fifty-five points, a total of five hundred five points.

After the test was first administered the papers were scored and handed back. One period—two hours—was spent discussing the results and giving the class instruction in the mathematics involved. All students wishing to have individual help were given opportunity to receive it. Half of the students took advantage of such help, which varied from very brief aid to clear up a cloudy point to extended programs of guided practice.

TABLE I.—SCORES OF FIFTY-TWO STATISTICS STUDENTS ON MATHEMATICS TEST

	Range in scores	Mean score	SD	Per cent correct
Test I	0–20	14.8	4.6	74.0
Test II	9–20	17.9	2.6	89.5

Table I summarizes the scores of the students on Test I (administered at beginning of course) and Test II (administered at end of course). The ratio of the difference between the means of Test I and Test II to its standard error is 4.09, indicating a significant or typical improvement. Three of the fifty-two students made slightly lower scores on the second test than on the first. These three made very high scores both times, however. All the rest, except those having perfect scores on both tests, made gains. The greatest gains were made by those below average on the first test; they made a mean gain of 5.8, as compared with a mean gain of 3.1 for the entire group. The greatest gains were made in skill in dealing with those processes which caused the students most trouble on the initial test, namely, the four fundamental processes with signed numbers and decimal fractions, percentage, and square root.

It would not be safe to conclude that the gains made were entirely due to the training given by the instructor. Constant practice throughout the course probably was an important factor in improvement. The chief contribution of the instructor's efforts at mathematical instruction were two-fold. They convinced the timid that the mathematics of elementary educational statistics were rather simple and not an insurmountable obstacle to success in the course. They called to the attention of the complacent their need to brush up on the simple arithmetic and algebra needed.

Correlation coefficients showing the inter-relationships of the various factors measured in this study are listed below:

Factors correlated	*r*	PE of *r*
Test I—Test II	+.61	±.06
Test I—total scores	+.45	±.08
Test II—total scores	+.53	±.07
Test I—examination scores	+.63	±.06
Test II—examination scores	+.57	±.06
Test I—problem scores	+.13	±.10
Test II—problem scores	+.25	±.09

It would seem that, in the case of these fifty-two students, there was a definite relation between ability in certain elementary mathematics and success in a course in educational statistics. The relation is much greater between mathematical ability and ability on examinations than between mathematical ability and ability to work problems.

No doubt this is due to the absence of the time factor when working the practice problems. Mathematical ability at the end of the course bore more relation to success at working problems than mathematical ability at the beginning of the course. The reverse was true of success on examinations.

SUMMARY

(1) A significant proportion of fifty-two students taking elementary educational statistics labored under definite disabilities in the mathematics needed for the course.

(2) Given a short period of training and plenty of practice material these students made significant gains in mathematical ability during one semester.

(3) A reasonable command of certain arithmetic and algebraic processes is a factor determining success in educational statistics, but by no means the sole factor.

The writer is convinced of the value of administering a brief test of the necessary mathematics to students in educational statistics at the outset of the course. Such a test serves both as a basis for remedial work with the few students who seriously need it and as a guide to the majority of the students, who need to review certain elementary mathematics in order to achieve the best possible success in statistics.

PREDICTING SUCCESS IN COLLEGE: A STUDY OF VARIOUS CRITERIA

SYLVESTER B. SCHMITZ

St. Benedict's College, Atchison, Kansas

For a number of years most colleges and universities have required freshmen to submit to a battery of entrance tests. The uses made of these tests differ, of course, with the different institutions. The writer has been connected with the administration of test batteries and became interested in their predictive value for success in college. This study accordingly represents a comparison of the several tests, singly and collectively, with the subsequent success of the students during one year or semester. The study likewise includes a comparison of the high-school marks with success in college.

The following tests were administered:

1. American Council Psychological Test.
2. Army Alpha Intelligence Test.
3. Iowa Reading Examination (Rate and Comprehension).
4. Purdue Placement Test.
5. Spelling Test (three hundred words prepared locally).

This study includes the tests administered to entering freshmen for the years 1934–1935 and 1935–1936. Success in college was determined by the Scholastic Quotient which was obtained as follows: The total number of quality points earned by a student during the year or semester was divided by the total number of hours of work carried. Quality points were awarded on the basis of the semester mark. For example, a three-hour course with grade A merited nine quality points; a three-hour course with grade C merited only three quality points; a three-hour course with grade F merited three negative quality points.

The Scholastic Quotients for the year 1934–1935 include the semester marks for the entire year, while those for the year 1935–1936 group include the first semester only. The high-school scholastic quotients for the entire group were determined in much the same manner.

Correlations between the various tests and the scholastic quotient were calculated by the Pearson Product Moment method. These correlations are listed in Table I.

TABLE I

	r	PE
1	.644	.075
2	.583	.033
3	.580	.034
4	.576	.034
5	.573	.034
6	.515	.037
7	.463	.040
8	.419	.053
9	.546	.037
10	.635	.031
11	.659	.030
12	.620	.032

Number 1 represents the correlation between the college quotient and the high-school quotient; (2) the correlation between the college quotient and the American Council Psychological Test; (3) between the college quotient and the total score on all tests; (4) between the college quotient and the Army Alpha; (5) between the college quotient and the Iowa Reading-comprehension; (6) between the college quotient and total reading, *i.e.* comprehension and rate; (7) between the college quotient and the Purdue Placement Test; (8) between the college quotient and spelling; (9) between the high-school quotient and the Army Alpha; (10) between the high-school quotient and the American Council Psychological Test; (11) between the high-school quotient and the total score on all tests; (12) between the high-school quotient and total reading score.

The first eight correlations in Table I may be considered criteria for determining success in college and accordingly are arranged in the order of their magnitude. Thus, the grades made by the student in high school appear to be the best criterion of success in college, having a correlation of 0.644, PE, .075, with the scholastic quotients in college. The American Council Psychological Test ranks second as a criterion of success having a correlation of 0.583. The total score on all tests takes third place as a criterion of success in college. The other correlations in the table are to be read in a similar manner. Our interest in this study naturally centers around the first eight of these correlations.

These correlations are all positive and significant although they are not large enough to be used for predictive purposes. Two facts in this table are noteworthy, first, that the high-school marks are after all the best single criterion of subsequent success, and, second, that Spelling should have such a large correlation as 0.419. A study of this table likewise reveals the fact that the American Council Psychological Test

is only slightly better than the Army Alpha as a criterion of success in college.

Since these correlations cannot conveniently be used for predictive purposes, another method of showing the relationship of each of the measures with success in college was employed. The method consists simply in using the quadrant scheme of comparing each measure with success in college. Graph I illustrates the method.

In this graph a comparison is made between the scholastic college quotients of the freshmen for the years 1934–1936 and their scores on the American Council Psychological Test. The numbers along the top of the chart from left to right represent, respectively, the lowest quartile (Q_1), the median, and the upper quartile (Q_3). Thus any student whose scholastic quotient in college was less than 1.51 would be listed in the small square in the uppermost left corner. A student whose scholastic quotient in college ranked between 1.51 and 2.07 would appear in the little square to the right of the square on the extreme left. Similarly, a student whose scholastic quotient ranked above 2.56 would appear in the small square at the top on the extreme right.

Similarly, the numbers at the left of the graph, namely, one hundred eight, one hundred fifty-one, and one hundred ninety-five, represent, respectively, the lowest quartile, the median, and the upper quartile on the American Council Psychological Test. Thus, there were twenty-five students (lowest square at left) whose scores on the test fell below one hundred eight; thirteen students whose scores on this test ranked between one hundred eight and one hundred fifty-one; seven scores ranking between one hundred fifty-one and one hundred ninety-five and one score above one hundred ninety-five. In determining a student's position on this graph, both his scholastic quotient and his score on the American Council Psychological Test had to be considered simultaneously. Thus, in the upper right square appears the number twenty-two. This means that twenty-two students of the entire group of one hundred eighty-four had scholastic quotients above 2.56 and made scores of one hundred ninety-five or better on the psychological test. Similarly, the number twenty-five in the lower left-hand corner means that twenty-five students had scholastic quotients below 1.51 and less than a score of one hundred eight on the psychological test.

The number thirty-five at the top outside the large square and under "Total" means that thirty-five students of the entire group

made scores on the psychological test of one hundred ninety-five or better. Of these thirty-five students, only twenty-two had scholastic quotients above 2.56, eight ranked between 2.07 and 2.56 in SQ, four ranked between 1.51 and 2.07, while one had a scholastic quotient below 1.51. (All scholastic quotients were increased by one unit to avoid dealing with negative numbers.) Similarly, the number fifty-seven at the right means that fifty-seven students made scores on the psychological test between one hundred fifty-one and one hundred ninety-five. The numbers forty-four and forty-eight at the right of the graph are read in a similar manner.

The numbers at the bottom of the graph from left to right, namely, forty-six, forty-four, forty-four and fifty, mean that forty-six students of the entire group have scholastic quotients below 1.51, forty-four below the median or average (2.07), forty-four between 2.07 and 2.56, and fifty above 2.56.

THE QUADRANT SCHEME

The four small squares in the upper right-hand corner constitute Quadrant I, the four small squares at the upper left constitute Quadrant II. Similarly, for Quadrants III and IV. A student whose scholastic quotient in college exceeded 2.07 and whose score on the psychological test at the same time exceeded one hundred fifty-one would appear in Quadrant I. A total of sixty-seven students appear in this quadrant, *i.e.* the sum of the numbers in the small squares. This means that sixty-seven students rated high or above average on both scholastic quotient and psychological test scores. These sixty-seven students actually measure up to expectation with regard to their college work. In other words, since their scores on the psychological test were above average at the beginning of their college work, we naturally expect them to do better than average work in college. Since these sixty-seven measured up to expectation we can consider Quadrant I the normal group. However, this quadrant contains only a part of the normal group. Quadrant III contains those students whose scores on the psychological test were below average (below one hundred fifty-one) and whose scholastic quotients at the same time were below average, *i.e.* below 2.07. Students with psychological scores below average will normally find college work more difficult, hence, we normally expect them to fall below the average in their college work. Quadrant III contains sixty-five cases, that is, the sum of the numbers in the small squares, twenty-five, seventeen, ten and thirteen. These again

measure up to expectation. Consequently both quadrants I and III constitute the total normal group.

On the other hand, Quadrants II and IV, contain those students whose college work does not coincide with expectancy. Thus, in Quadrant II we find twenty-five cases in which the scores on the psychological test were above average but the college work was actually below average. Quadrant IV contains those subjects whose scores on the psychological test were below average (below one hundred fifty-one) but who by reason of exceptional application and study actually ranked above average in college work. Quadrant II, therefore, would contain those students who on account of lack of application, sickness, or other factors, do not measure up to expectation. These are the problem cases. In Quadrant IV would be listed those students who by reason of exceptional industry exceed expectation. These two quadrants, therefore, will be considered the abnormal group.

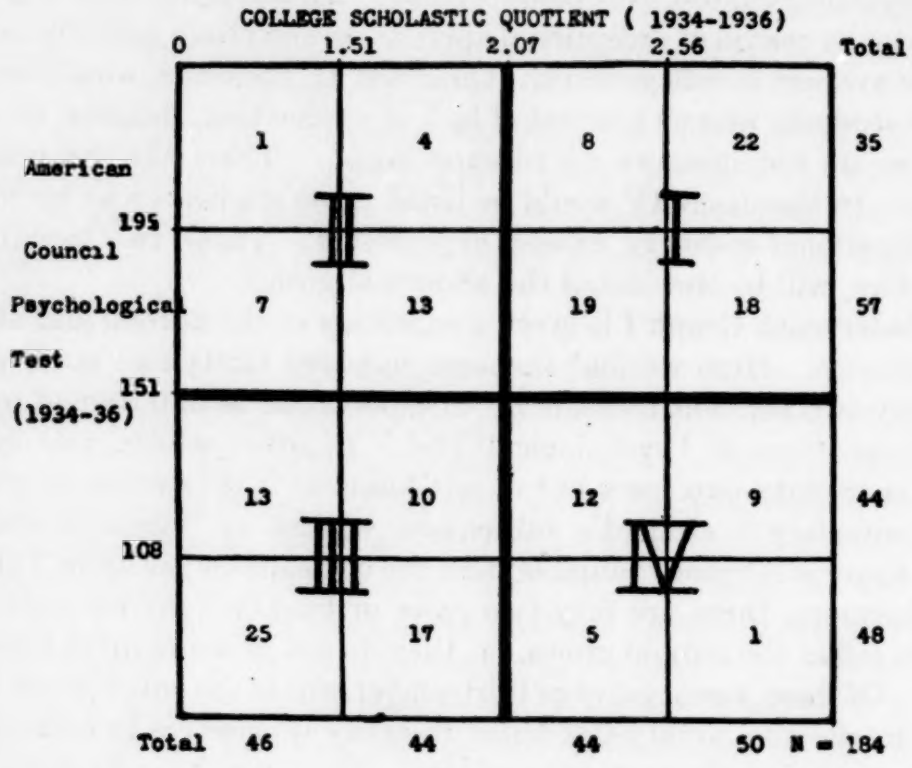
Underneath Graph I is given a summary of the normal and abnormal groups. Here we find that one hundred thirty-two students or seventy-two per cent measure up to expectation as determined by the American Council Psychological Test. In other words, this means that in seventy-two cases out of one hundred it is possible to predict approximately a student's subsequent success or failure in college. This figure is far more valuable than the correlations set up in Table I. Furthermore, there are fifty-two cases or twenty-eight per cent that do not fall in the normal group, *i.e.* they do not measure up to expectation. Of these, twenty-five or thirteen per cent of the entire group have good intellectual ability but failed to apply themselves in college and thus rated below the average in college. These are the individuals who need guidance, advice, and probably some prodding at various times during their first semester.

Other interesting information can be obtained by a careful study of this graph. Consider the numbers appearing in the small squares along the bottom of the graph, namely, twenty-five, seventeen, five and one, a total of forty-eight cases. These individuals ranked in the lowest quartile on the psychological test. Note that twenty-five of these forty-eight cases likewise rated in the lowest quartile in college work, *i.e.* to the left of the vertical line drawn through 1.51 at the top of the page. In other words, it is possible to predict in fifty-two cases out of one hundred (twenty-five out of forty-eight) the type of college work a student will do in college if we know in advance that his score on the psychological test is in the lowest twenty-five per cent of the group.

Keeping in mind that all scholastic quotients have been increased by one unit, these students would actually have scholastic quotients below 0.51. Approximately half of their semester marks would be "*D*" or less.

Again, a student whose score on the psychological test is below one hundred eight has only one chance in eight of doing average college

GRAPH I



Quadrant I = 67
Quadrant III = 65
Normal = 132 = 72%

Quadrant II = 25
Quadrant IV = 27
Abnormal = 52 = 28%

work or better. There is only one chance in forty-eight that a student in this group will appear in the highest quartile in college work. Furthermore, there are forty-two chances in forty-eight, or nearly eighty-seven in one hundred, that a student whose score on the psychological test is below one hundred eight will not do average college work. This is perhaps the most important information yielded by this study. By far the majority of the individuals rating low on the psychological test have very little hope of success in college.

Now, consider the numbers appearing in the small squares along the top of the graph, namely, one, four, eight and twenty-two, making a total of thirty-five. There are twenty-two chances in thirty-five, or nearly sixty-three in one hundred, that a student ranking in the upper quartile on the psychological test, will likewise rate in the upper quartile of his group in college work. A student in this group has thirty chances (twenty-two plus eight) in thirty-five, or nearly eighty-six in one hundred, of getting better than average rating on his college work. However, there is only one chance in thirty-five that his scholastic rating will be in the lowest quartile, and only one chance in seven that his rating in college will be below the average, (five in thirty-five).

In a similar manner a comparison was made between the scholastic quotient and (1) total score on all tests previously listed, (2) high-school scholastic quotient, (3) the Army Alpha Intelligence Test, and (4) the Iowa Reading Test. The results of these comparisons are listed by quadrants in Table II.

TABLE II

Criteria compared	Normal				Abnormal				Total cases
	I No.	III No.	Total	Per cent	II No.	IV No.	Total	Per cent	
College scholastic quotient and American Council Psychological Test	67	65	132	72	25	27	52	28	184
College scholastic quotient and total score —all tests	64	68	132	72	22	30	52	28	184
College scholastic quotient and high-school quotient	67	67	134	75	19	25	44	25	178
College scholastic quotient and Army Alpha Intelligence Test	72	62	134	71	28	28	56	29	190
College scholastic quotient and Iowa Reading Test	68	67	135	72	23	28	51	28	186

Table II read as follows: sixty-seven cases are found in the first quadrant on both scholastic quotient and the American Council Psychological Test, and sixty-five cases in the third quadrant, making a

total of one hundred thirty-two cases that measure up to expectation and, therefore, constitute the Normal Group. For these it is possible to predict success in seventy-two per cent of the cases. Similarly, continuing on the first line, we find twenty-five cases in the second quadrant, twenty-seven in the fourth, making a total of fifty-two in the Abnormal Group. These do not measure up to expectation. The last column gives the total number of cases included in the comparisons. The other lines are read in the same manner.

Perhaps the most remarkable fact disclosed by Table II is the similarity of the percentages of the normal groups as determined by the several comparisons, three of the five being identical. (See column 4.) The high-school scholastic quotient appears to have the highest predictive value, since the comparison shows that seventy-five per cent of the cases measure up to expectation. On the other hand, the Army Alpha is slightly inferior to the other creteria used in this study, since it yields only seventy-one per cent of the cases that measure up to expectation.

In spite of the high degree of similarity of the various instruments for predicting success in college, the writer is of the opinion that it would not be prudent and safe on the part of an administrator to use one instrument alone. The writer observed in the study of many cases in this group that did not measure up to expectation, that an individual might have a high rating in high-school scholastic quotient and a comparatively low rating on the American Psychological Test or on the total score on all tests. On the other hand, it was noted that several individuals made a comparatively low rating in high-school scholastic quotient, with average score on the intelligence tests, and their subsequent work in college was average or better. In these latter cases it appeared quite evident that the individual attention these students received in college together with an environment more conducive to study than that to which they had been accustomed in high school were the chief factors that enabled them to do satisfactory college work. It thus appears that it is safest to use several criteria since one instrument serves as a check on the others.

SUMMARY

This study gives the comparative value of several criteria for predicting success in college. Success in college was measured by the individual scholastic quotient, *i.e.* the total number of quality points divided by the total number of hours of work carried. Com-

parisons were made between the scholastic quotient and (1) the American Psychological Test, (2) the Army Alpha Intelligence Test, (3) Iowa Reading Test, (4) total score on all tests, and (5) the high-school scholastic quotient.

Two methods of determining the value of each instrument were employed, namely, the coefficient of correlation, and the quadrant scheme. Both methods agree in showing that the high-school scholastic quotient is the most efficient single instrument for predicting success. The two methods do not agree with regard to the value of the other instruments, although the discrepancies are very slight with the exception of the Army Alpha. Thus, the quadrant scheme shows that the Army Alpha is inferior to the other instruments, whereas the coefficient of correlation for this test is .576 as compared with only .515 for the Iowa Reading Test. In general it might be stated that these instruments have about equal value. An administrator will, however, be on the safer side, if he employs several criteria, because one will serve as a check on the other.

THE CURVE OF FORGETTING AND THE DISTRIBUTION OF PRACTICE

HOWARD EASLEY

Duke University

It is quite generally observed and well attested by experimental findings that distributed practice produces greater results than does concentrated practice. This is true despite the fact that until very recently the law of disuse was universally accepted and respected. If we are to judge by the discussions of the two principles separately, the inconsistency involved has not always been recognized. On the other hand, its recognition in some quarters may have occasioned some of the attempts to account for the effects of distributed practice. One prominent explanation given is in terms of the curve of forgetting. The reasoning is as follows: The curve of forgetting shows (and most published curves do seem to show this) that we forget most rapidly *at first* and more slowly later. Therefore, practice periods should be close together *at first* to eliminate this rapid forgetting, and more frequent later on when forgetting is slower. Now obviously the term *at first* is used here to refer to two totally different periods, but the reasoning implies that it refers to the same early period. When we say that forgetting takes place rapidly *at first* we mean in the early period after practice ceases. When we say that practice periods should be close together *at first* we mean in the early period of learning.

It is true that most of the experimental attacks on the problem of distribution have been concerned merely with the relative merits of various degrees of distribution, and not with various modes of distribution. Gordon,[3] Pechstein,[5] Reed,[6] and Ulrich[8] have made studies which illustrate this type. If we make a practical application of the Ebbinghaus curve of forgetting and say that we should eliminate the period of rapid forgetting, then, since that period comes first, we cannot eliminate it without also eliminating all later periods. Consequently practice periods should necessarily be as close together as possible. In other words, we should have complete concentration of practice. Yet the fact remains that the great majority of studies on this point show results in favor of distributed practice. It was this inconsistency that led the writer to question the validity of the Ebbinghaus curve.[1]

A few studies have included the problem of the mode of distribution. McClatchy[4] found that in maze learning the introduction of a twenty-

four-hour interval gave better results in the early trials (third to fifth), whereas if a forty-eight-hour interval was to be introduced it should come later (after the seventh trial). Tsai[7] compared the results of three modes of distribution, viz. increasing intervals, regular intervals, and decreasing intervals (called respectively negatively accelerated, linear, and positively accelerated, to match three possible forgetting curves). The increasing intervals were found to give best results both in original learning and in review. This was explained entirely in terms of the Ebbinghaus curve.

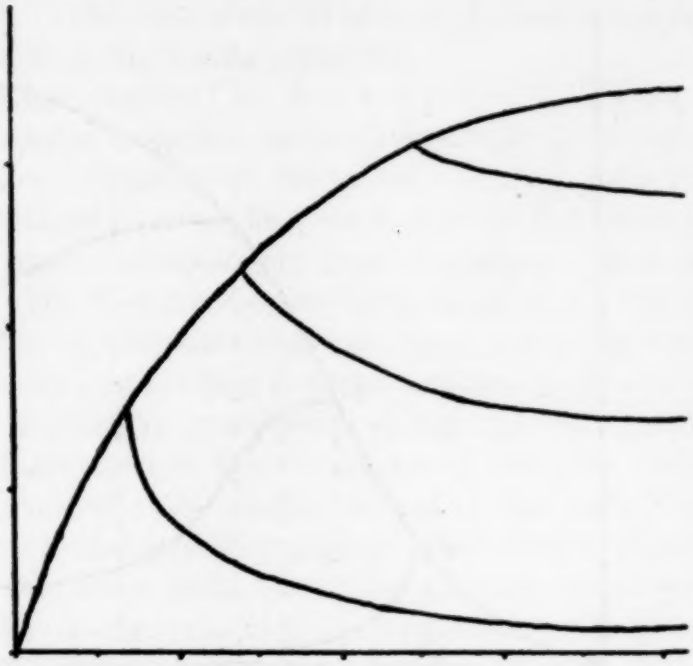


Fig. 1. Relation between level of learning and rate of forgetting (after Gates).

Another type of explanation might be inferred from a set of relations proposed by Gates.[2] These relations are illustrated in Fig. 1. While the forgetting curve here is of the Ebbinghaus type, with higher levels of learning the initial period of rapid forgetting is less marked, and the total amount forgotten is less. It would follow that with higher levels of learning, longer intervals may be introduced without so much loss. This explanation has much to recommend it in preference to any one based simply on the nature of the forgetting curve. But it has certain limitations. It merely accounts for the decreasing undesirability of long intervals. Yet at all levels of learning the most rapid forgetting is immediately after practice ceases. We still have to

account for the actual desirability of the longer intervals, as shown by McClatchy[4] and Tsai,[7] or for the desirability of distribution at all, as shown by practically all investigations of the subject.

It may be suggested that fatigue is the determining element. Possibly at best there is an inevitable loss from disuse and intervals are to be suffered in any event, but it is better to suffer a certain amount of forgetting than the effects of boredom or fatigue. This, however, is inconsistent with two types of observations. In the first place, every experimenter has probably observed incidentally what James described

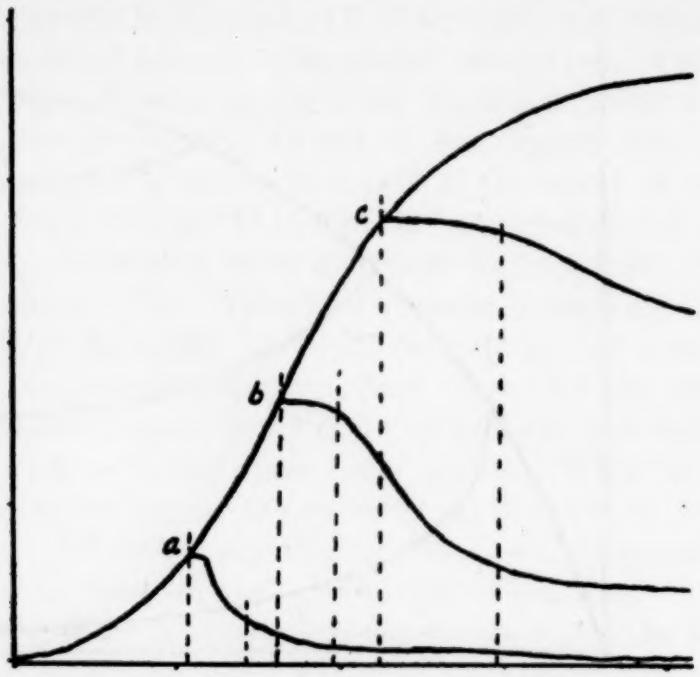


Fig. 2. Relation between level of learning and rate of forgetting.

as "learning to swim in Winter and to skate in Summer." In the second place, we find distributed practice superior in cases where the practice periods are too short to account reasonably for boredom or fatigue, and where the rest intervals are far longer than necessary to recover from these effects, even if they were present. The experiments of Gordon[3] and Ulrich[8] are notable examples of this. It hardly seems probable that three readings of a short selection of prose is too much for a college student, or that one trip through a maze is sufficient to exhaust a rat. In any event it should seem that a twenty-four-hour rest period would be enough to permit recovery from such tasks. But longer intervals produced better results.

A modification of the relations suggested by Gates, combined with a consideration of certain characteristics of complex learning functions, seems to give a more consistent explanation of all the facts concerning this problem. The writer has suggested elsewhere[1] that the true curve of forgetting is probably positively accelerated at first. The relation between the degree of learning and the rate of forgetting may then be represented as in Fig. 2. Here the initial period of slow forgetting increases in length with higher levels of learning. The intervals between practice periods may then be increased without appreciable loss from disuse. This fact alone would not, however, account for the actual desirability of increasing intervals.

In any complex learning function, and probably in any simple one, we have in the total behavior pattern some right (to-be-learned) and some wrong (to-be-eliminated) responses. At any time the level of learning, as measured by a test, is a function of the relative numbers and relative prominence of these two types of responses. There is evidence to support the view that practice on the so-called wrong responses tends to fix them as it does with the right responses. Learning, then, involves fixing both the right and wrong responses, although we observe it only as increasing the relative prominence of the right ones, and therefore refer to the elimination of the wrong ones. In pure trial-and-error learning the conditions are usually such that the right responses will inevitably occur more frequently than any given wrong response. As a result we have the more rapid learning of the right responses.

It is true that in more rational learning some key response (usually a mental one) becomes well fixed by a small number of repetitions and thereby causes the elimination of large numbers of practiced responses to which it is completely antagonistic. Here the situation is complicated by complex patterns of well-learned rational behavior which defy attempts at accurate and detailed description. But it is probable that the basic principles are not different from those involved in trial-and-error learning.

Returning to the trial-and-error learning problem, let us suppose that at a given time a wrong response has been learned to an extent represented by *a* on the learning curve, Fig. 2, and a right response is learned to an extent represented by position *b*. At this point the introduction of as long an interval as can be tolerated without serious loss to the right response has a marked effect on the wrong one. At a later stage in the learning process when the right response has reached point *c* the interval not only may but should be lengthened to allow as

much forgetting as possible in the wrong responses without seriously affecting the right ones.

These relations are evident in problems like maze learning where we can actually observe the subject practicing the errors, *i.e.*, going into blind alleys. But in problems like learning a series of nonsense syllaables or committing a passage to memory they are not quite so evident. It should be noted that entering blind alleys, or stopping, or returning in the true path are not the only errors possible in maze learning, although they are usually the only ones recorded. Indecision, wondering, confusion, etc. are patterns of wrong responses, even though the right path may eventually be chosen. Such patterns are the "errors" practiced during memorizing. Even "trying" it is to be eliminated, for when the material is well-learned we are supposed to produce it without effort. We are frequently embarrassed by trying to produce a fact, a name, or a similar response. Later it "pops up" without our trying. We need, then, intervals in practice not only to rest but to forget these wrong responses, and with higher levels of learning these intervals may be lengthened without forgetting the right responses.

REFERENCES

1. Easley, H.: "An Analysis of the Curves of Learning and Forgetting Code Material." *J. Educ. Psychol.*, Vol. XXIV, 1933, pp. 634–640.
2. Gates, A. I.: *Psychology for Students of Education*, p. 311.
3. Gordon, K.: "Class Results with Spaced and Unspaced Memorizing." *J. Exper. Psychol.*, Vol. VIII, 1925, pp. 337–343.
4. McClatchy, V.: "The Optimal Position of a Rest Period in Learning." *J. Exper. Psychol.*, Vol. VIII, 1925, pp. 251–277.
5. Pechstein, L. A.: "Massed *vs.* Distributed Effort in Learning." *J. Educ. Psychol.*, Vol. XII, 1921, pp. 92–97.
6. Reed, H. B.: "Distributed Practice in Arithmetic." *J. Educ. Psychol.*, Vol. XV, 1924, pp. 248–250.
7. Tsai, L. S.: "The Relation of Retention to the Distribution of Relearning." *J. Exper. Psychol.*, Vol. X, 1927, pp. 30–39.
8. Ulrich, J. L.: *Distribution of Effort in Learning in the White Rat.* Behav. Monog., Vol. II, 1915, No. 10.

BOOK REVIEWS

GLENN D. HIGGINSON. *Psychology.* New York: The Macmillan Company, 1936, XIV + 646.

This *Psychology* is a distinctive and unorthodox volume that attempts, in the author's words, a "consistent and systematic psychology" as "the naturalistic relations intimately sustained between a functioning organism (man) and the sources of energy in his environment." Emphasis, originating from the *Gestalt* psychology, is placed upon man as a total organism, and synthetic explanations preferred to analytic. The functional properties of the whole man are to be understood in terms of their psychological products. These products—thoughts, perceptions, actions—emerge as a result of the interrelationship between man and his surroundings.

The excellence of this conception of psychology will be recognized clearly by the many workers who regard the human being as an entire, dynamic, outreaching organism, and not as a bundle of analyzable reflexes. It is with some disappointment, however, that one reads more thoroughly in the present volume. The author misses many opportunities to develop his thesis convincingly. For example, the treatment of "original nature" follows an instinct theory drawn largely from Darwin, Huxley, and James. The significant studies of the origins of traits following the line of Coghill's researches are ignored entirely. At the end of the volume, where the subject of personality should richly summarize the relation of the whole man to his environment, Higginson has only two topics under the heading of "Explanation." These are "endocrinology" and "psycholoanalytical"—surely a poor yield from so rich an ore.

The text contains a number of other faults, some of which are not minor. There are several serious failures of internal consistency. On one page it is stated that "inherent" characteristics will be regarded as "instinctive." Two pages later a principal caption reads "Some Non-Instinctive Forms of Behavior Which Have an Inherent Basis," under which "play," "collecting," and "work" are described. In a later chapter a heading interests the reader by announcing a consideration of "likenesses and differences between learning and memory." But alas, the paragraph contains little on this topic, being almost entirely devoted to the problem of instinctive *versus* acquired reactions.

Higginson's *Psychology* is provocatively stimulating at some points and simply exasperating at others. All psychologists should read it,

but the reviewer doubts that it will convey to the beginning student more understanding than confusion. LAURANCE F. SHAFFER.

Carnegie Institute of Technology.

WILLIAM S. SADLER. *Theory and Practice of Psychiatry.* St. Louis: C. V. Mosby Co., 1936, pp. XXII + 1231.

Dr. Dadler says that his motive is "to make this book of help to professional readers of all classes rather than to confine it to the recording of our experience for the exclusive perusal of the psychiatric specialist." To this end he is more particularly concerned with "the non-institutional management of nervous, emotional, and personality disorders. . . . " Such a work should be of interest and value to teachers, psychologists, and other non-medical professional group.

In the first two parts, containing some four hundred twenty-five pages, Saddler discusses theory of psychiatry and of personality. Unfortunately, this discussion is a very poorly integrated eclecticism. The author refers frequently to psychological literature but seldom evidences critical evaluation. On page 13 he says Freud rejected materialism, but on page 17 he speaks of "the materialistic basis of his [Freud's] teaching." He is a dogmatic proponent of sterilization as a panacea for social ills (*cf.* page 157), and appears totally unaware of contrary arguments or evidence. We may quote the definition of personality as a fine example of the loose thinking which is generally in evidence throughout the book. This definition reads as follows: "We might define personality as a self-conscious individual functioning in the presence of his fellows, and regard it as the sum total of the estimate of his fellows respecting those component factors which taken together, constitute personality." In the strictly clinical portions of the book there are some good cases, but even here the interpretations evidence the same loose logic found in the theoretical portions.

It is not our place here to speculate on the value of this book to members of the medical profession. For psychologists it has practically no value. For teachers, with their usually limited amount of psychological training, it would be very confusing and might even be harmful. The author has fallen far short of the ideal which he apparently set for himself. C. M. LOUTTIT.

Indiana University.